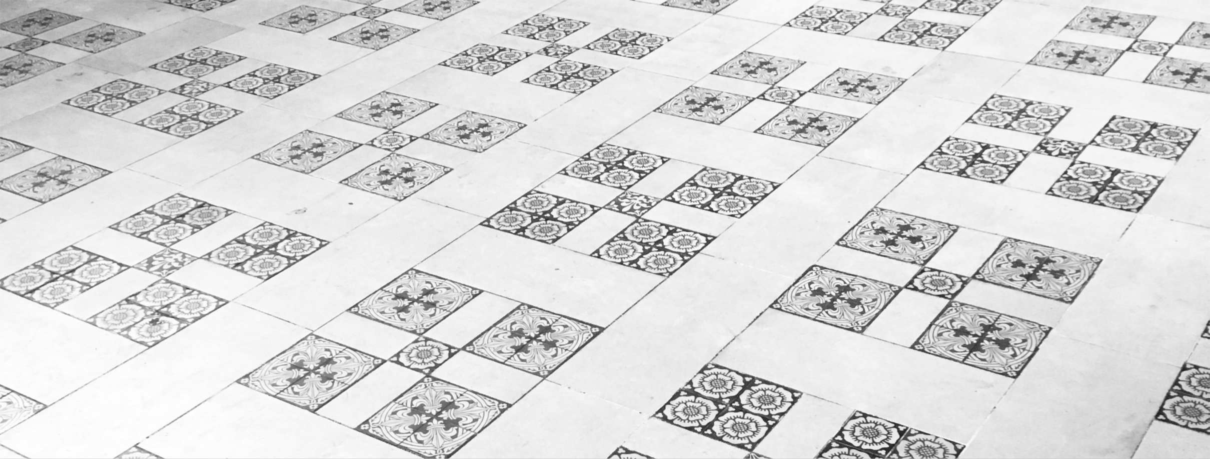
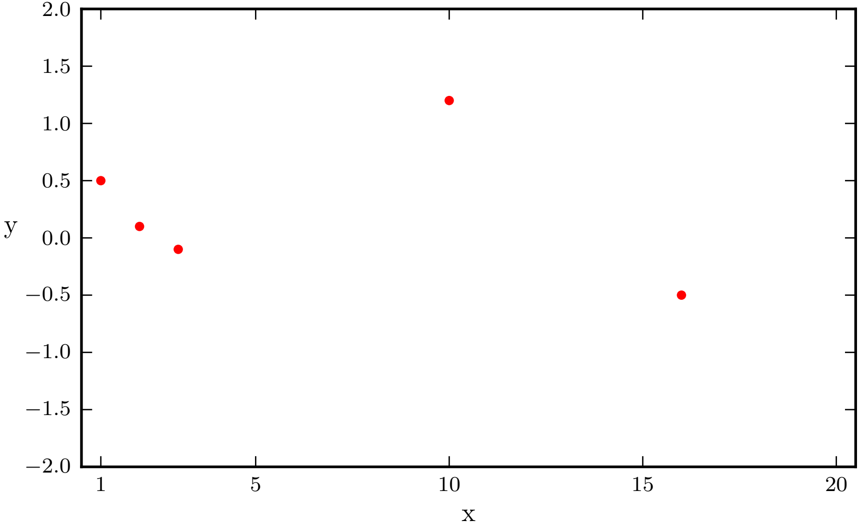


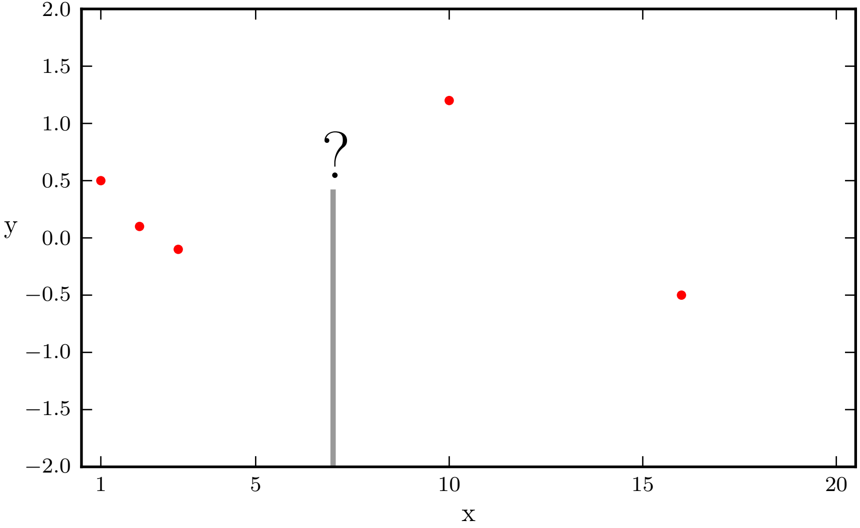
An Introduction to Gaussian Processes

Richard E. Turner
University of Cambridge

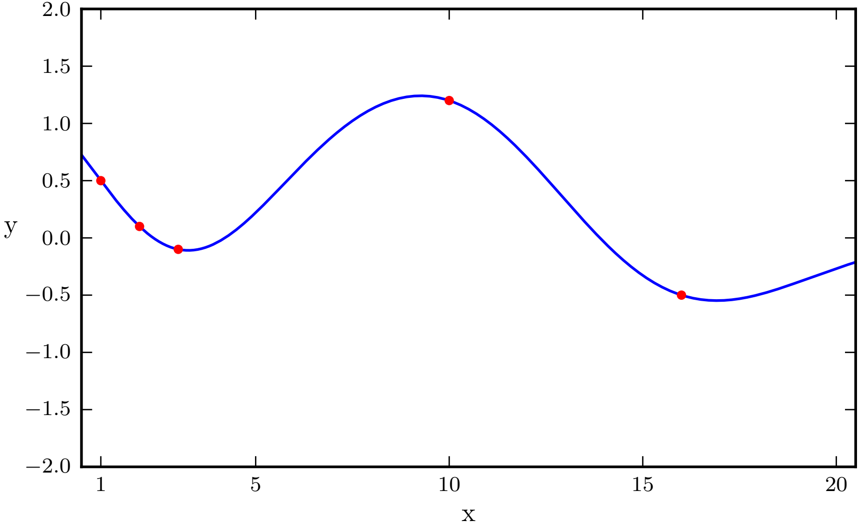
Motivation: non-linear regression



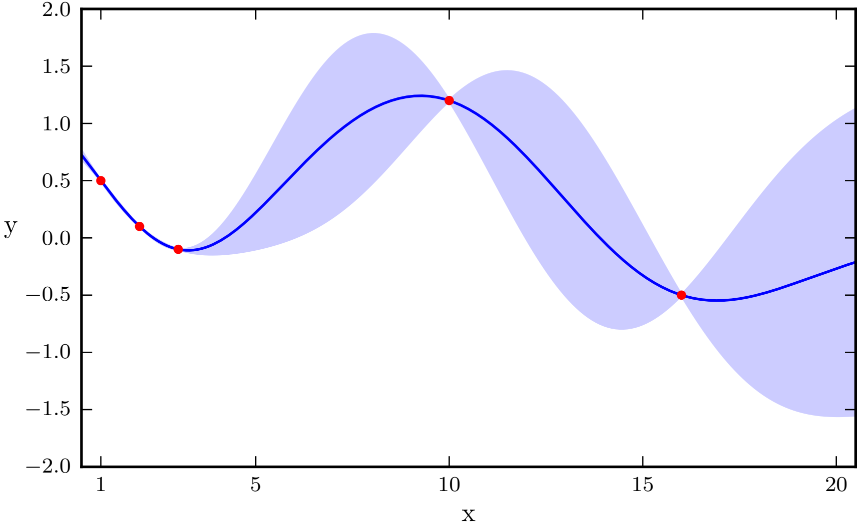
Motivation: non-linear regression



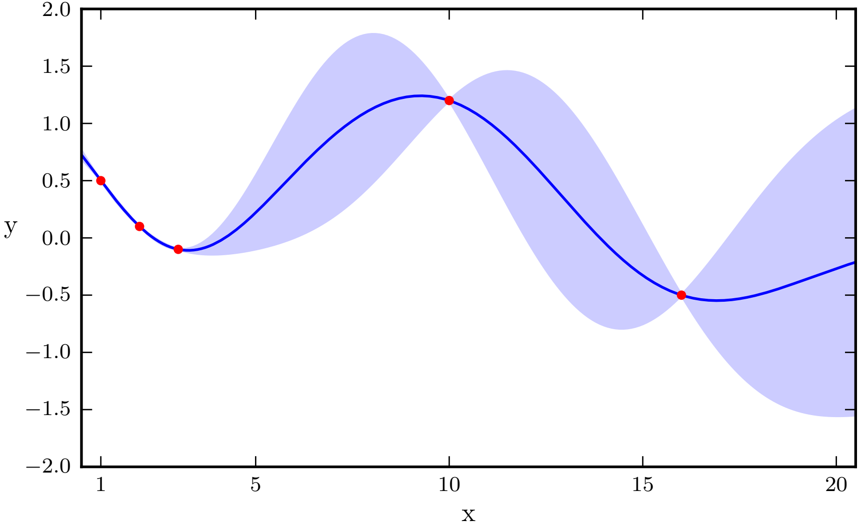
Motivation: non-linear regression



Motivation: non-linear regression

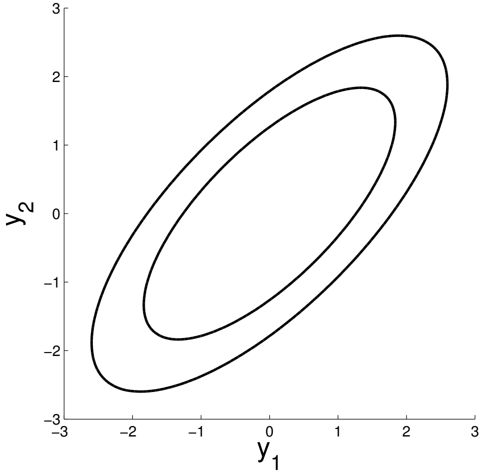


Motivation: non-linear regression. Can we do this with a plain old Gaussian?



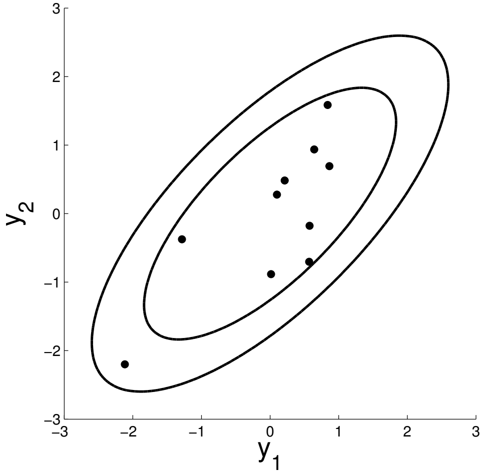
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top \Sigma^{-1} \mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



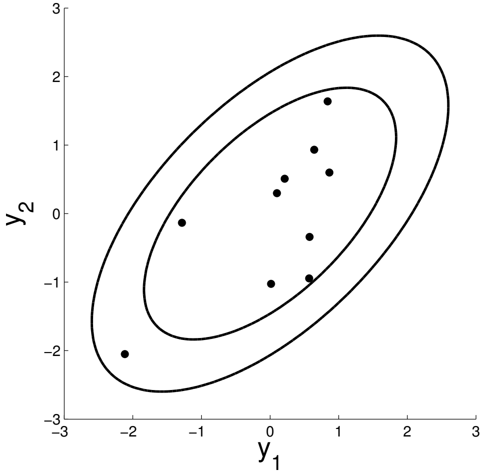
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



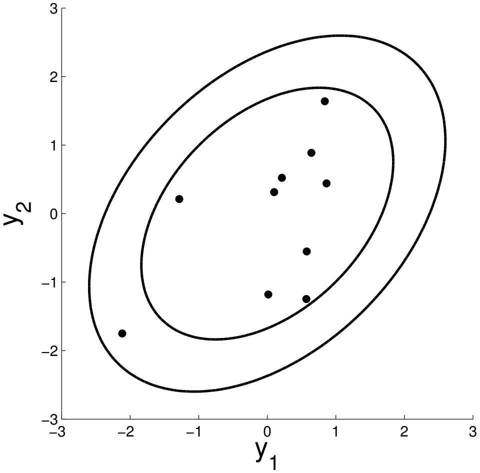
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top \Sigma^{-1} \mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .6 \\ .6 & 1 \end{bmatrix}$$



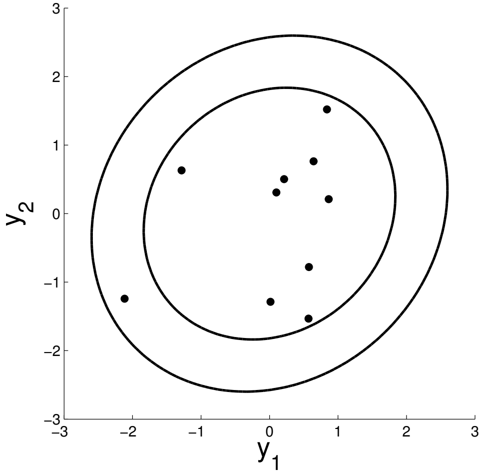
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .4 \\ .4 & 1 \end{bmatrix}$$



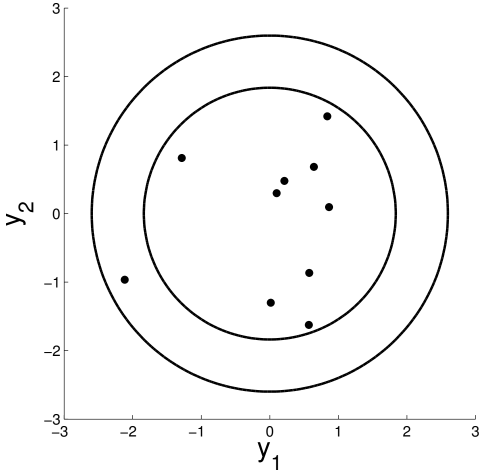
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .1 \\ .1 & 1 \end{bmatrix}$$



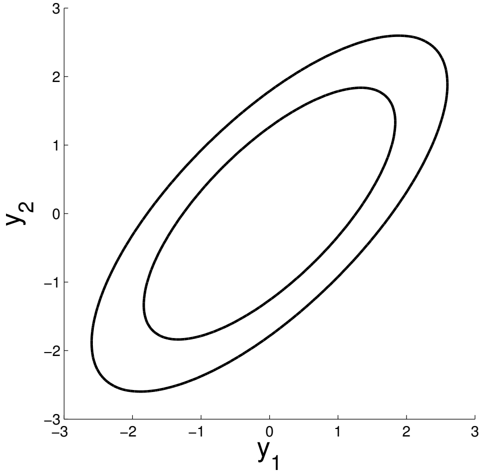
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



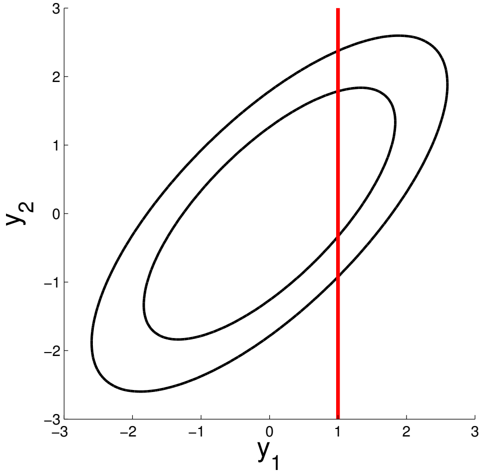
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\mathsf{T}\Sigma^{-1}\mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



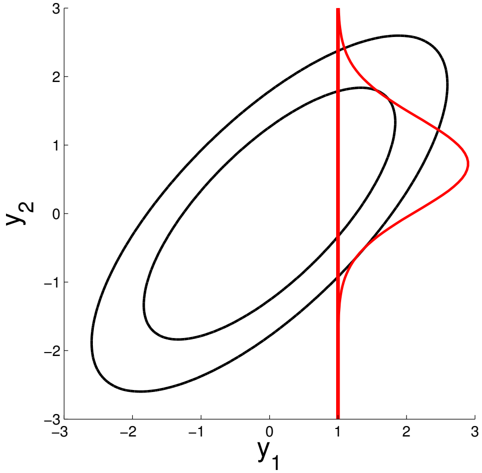
Gaussian distribution

$$p(\mathbf{y}|\Sigma) \propto \exp\left(-\frac{1}{2}\mathbf{y}^\top \Sigma^{-1} \mathbf{y}\right) \qquad \Sigma = \begin{bmatrix} 1 & .7 \\ .7 & 1 \end{bmatrix}$$



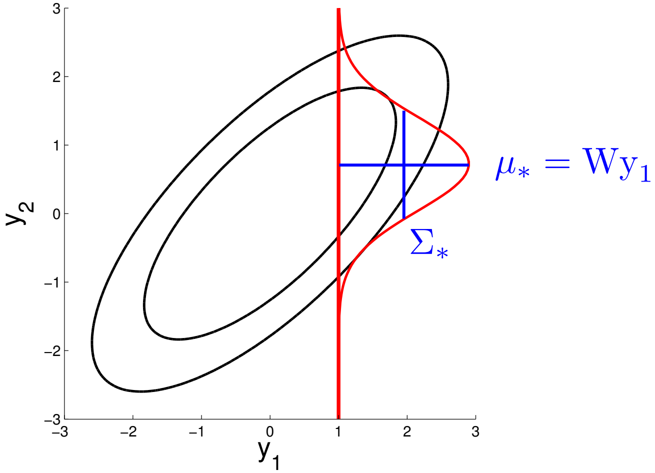
Gaussian distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



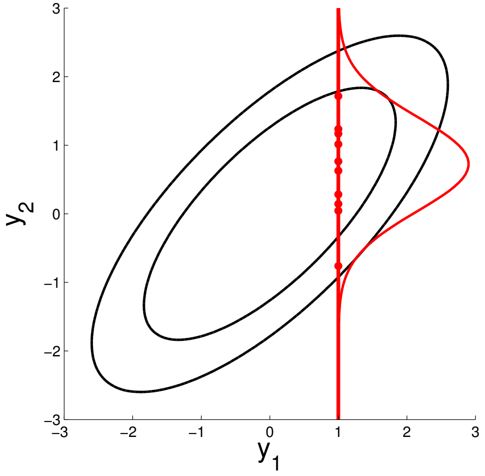
Gaussian distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



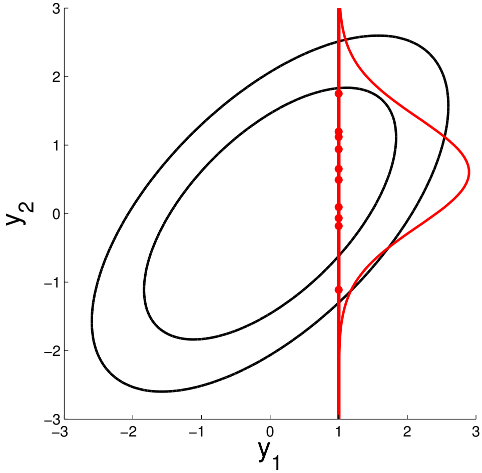
Gaussian distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$



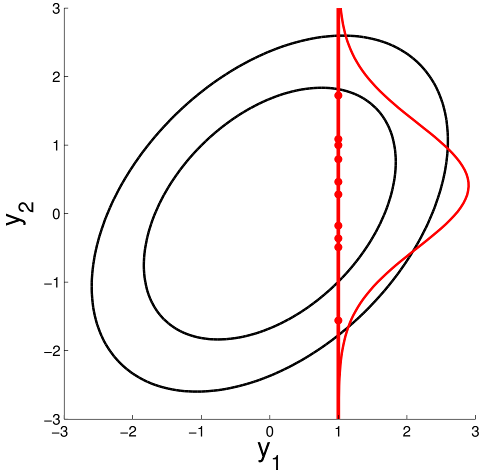
Gaussian distribution

$$p(y_2|y_1, \Sigma) \propto \exp\left(-\frac{1}{2}(y_2 - \mu_*)\Sigma_*{}^{-1}(y_2 - \mu_*)\right)$$



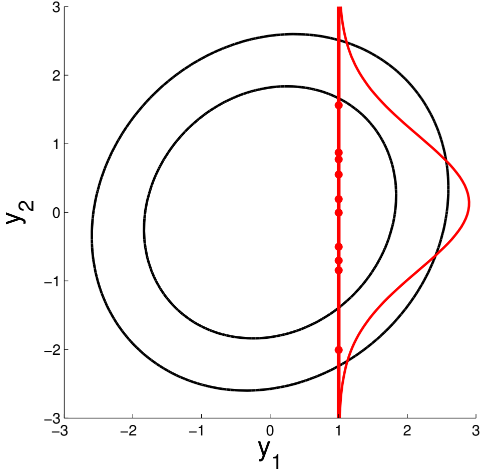
Gaussian distribution

$$p(y_2|y_1, \Sigma) \propto \exp\left(-\frac{1}{2}(y_2 - \mu_*)\Sigma_*{}^{-1}(y_2 - \mu_*)\right)$$



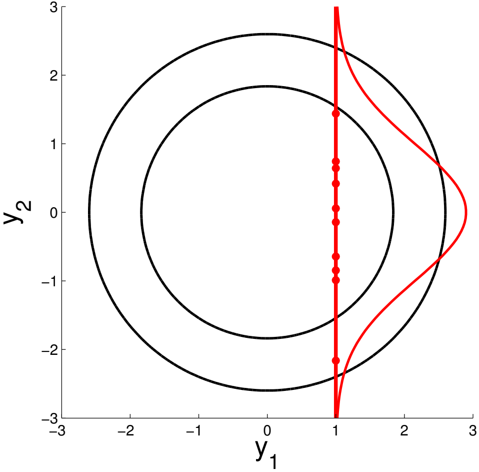
Gaussian distribution

$$p(\mathrm{y}_2|\mathrm{y}_1, \Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2 - \mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2 - \mu_*)\right)$$

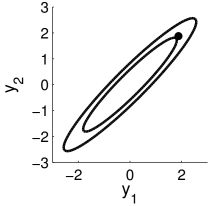


Gaussian distribution

$$p(\mathrm{y}_2|\mathrm{y}_1,\Sigma) \propto \exp\left(-\frac{1}{2}(\mathrm{y}_2-\mu_*)\Sigma_*{}^{-1}(\mathrm{y}_2-\mu_*)\right)$$

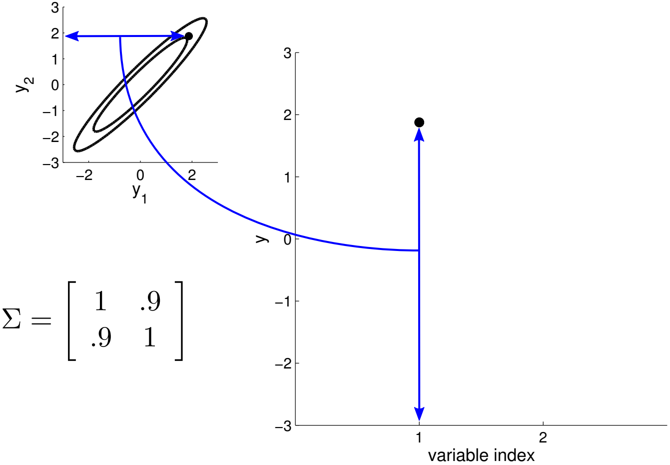


New visualisation



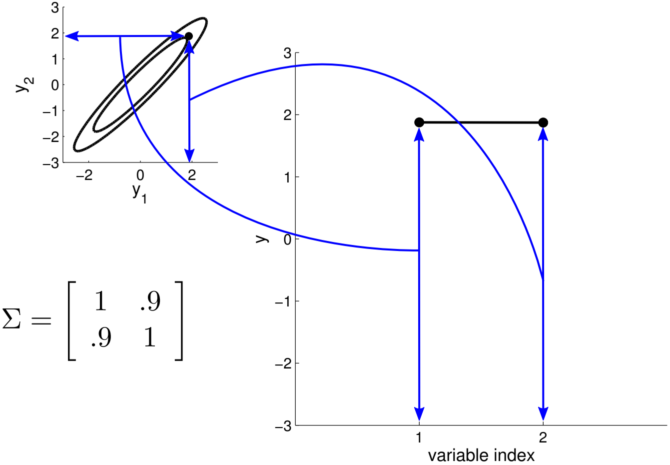
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation



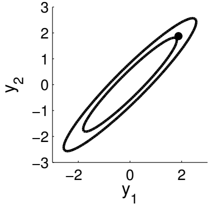
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

New visualisation

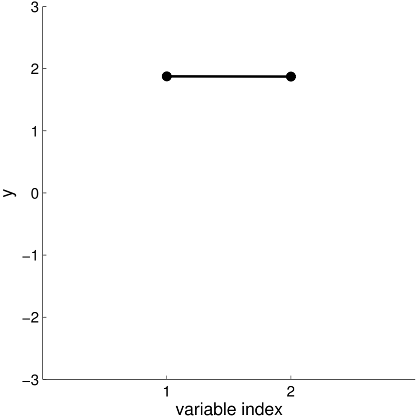


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

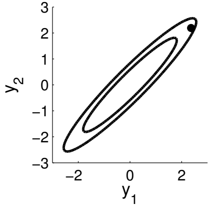
New visualisation



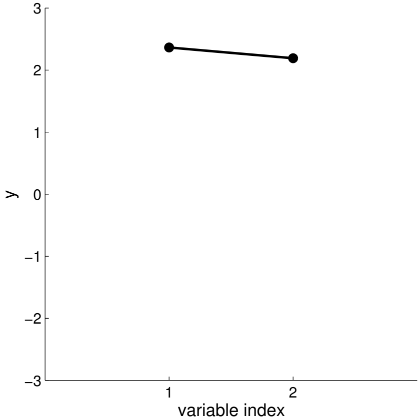
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



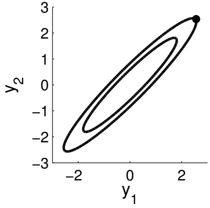
New visualisation



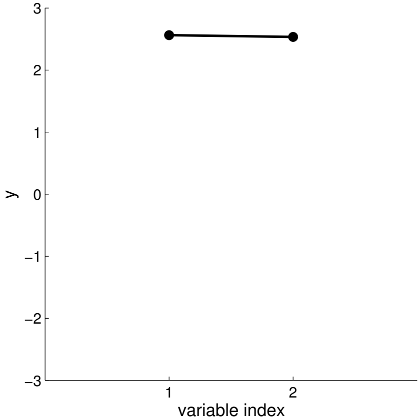
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



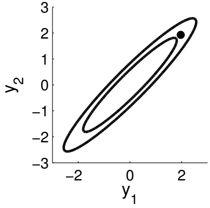
New visualisation



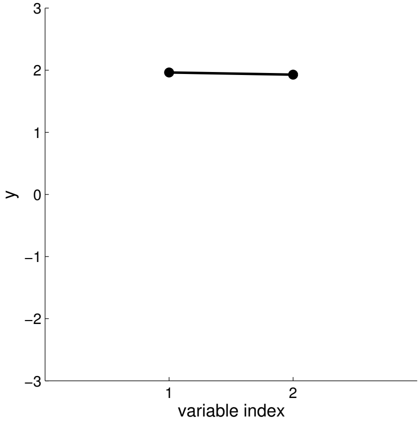
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



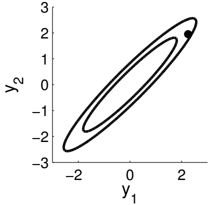
New visualisation



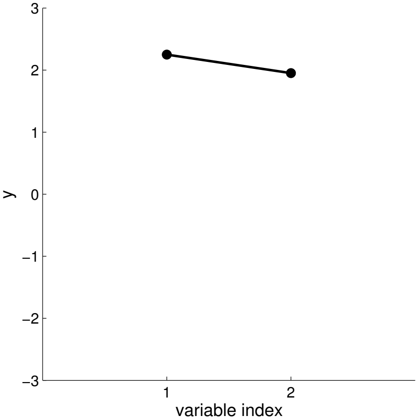
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



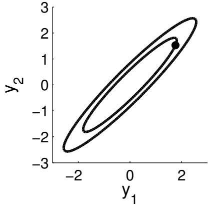
New visualisation



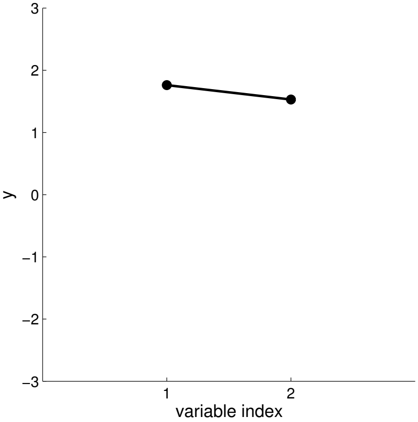
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



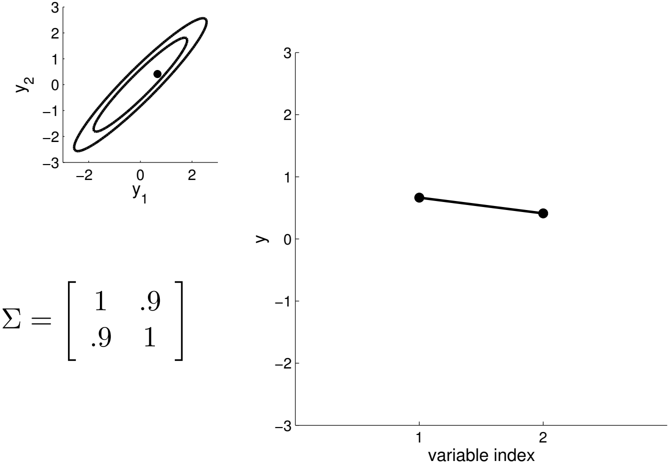
New visualisation



$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

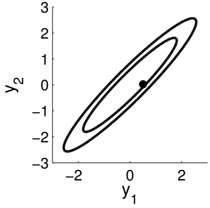


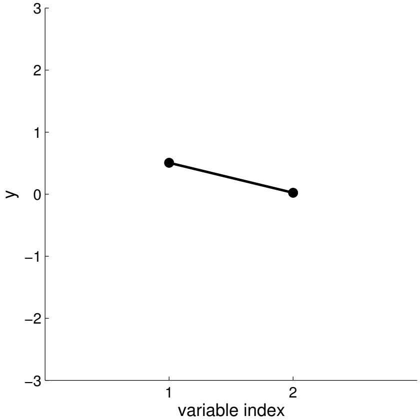
New visualisation



$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

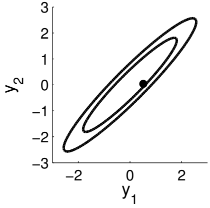
New visualisation



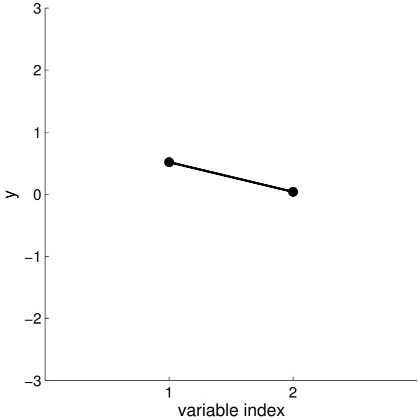


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

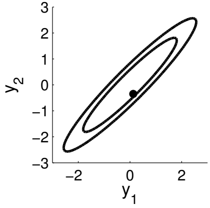
New visualisation



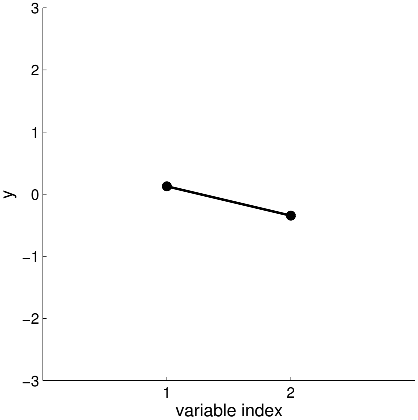
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



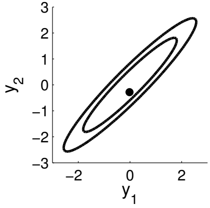
New visualisation



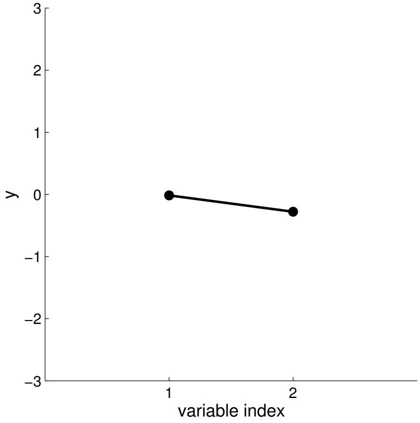
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



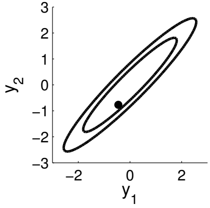
New visualisation



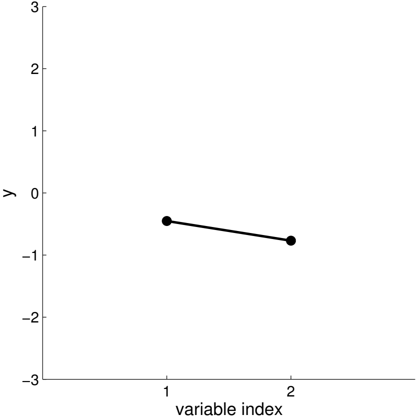
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



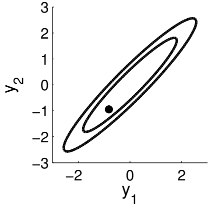
New visualisation



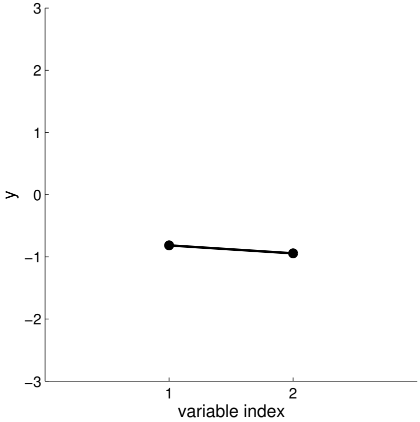
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



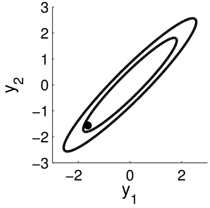
New visualisation



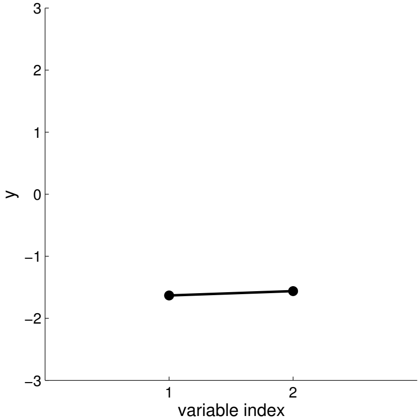
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



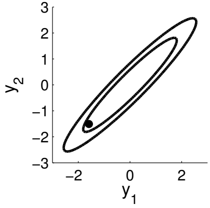
New visualisation



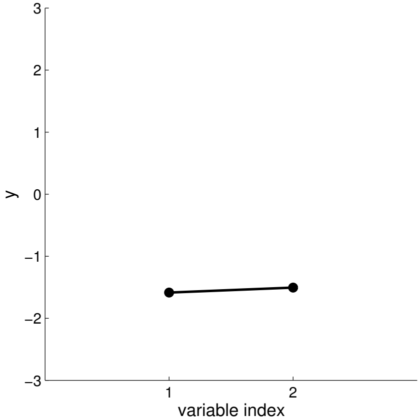
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



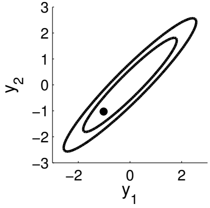
New visualisation



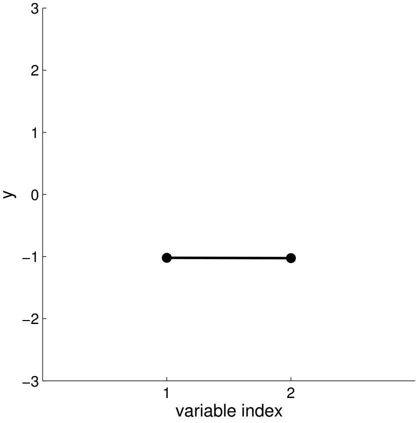
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



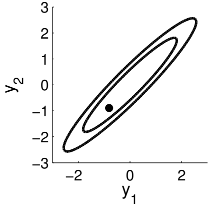
New visualisation



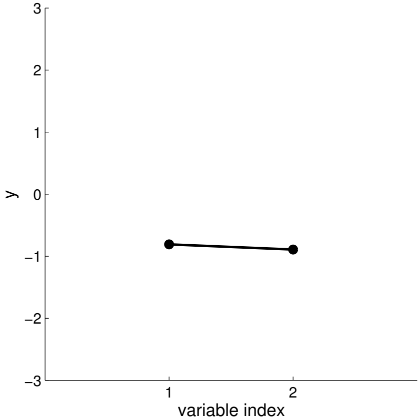
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



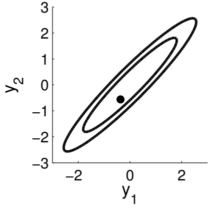
New visualisation



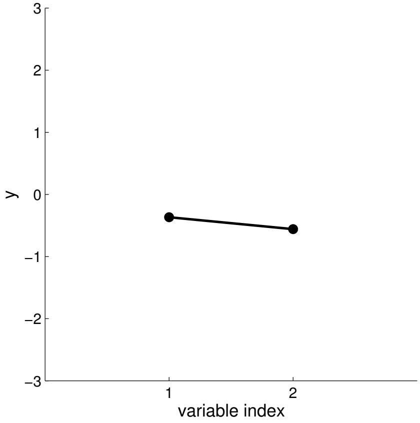
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



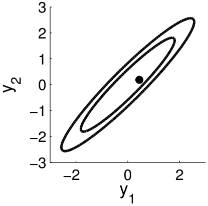
New visualisation

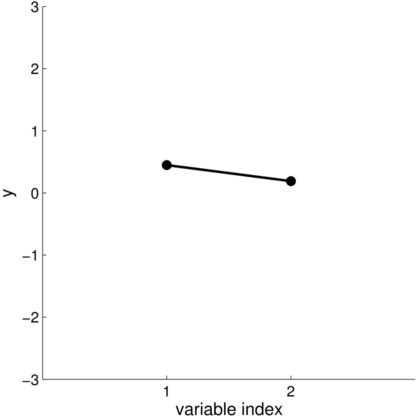


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



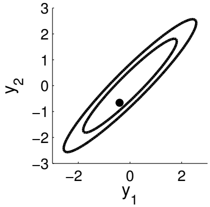
New visualisation



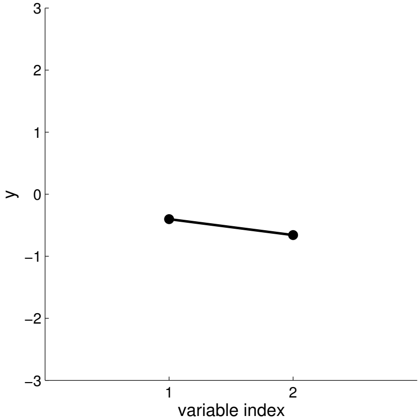


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

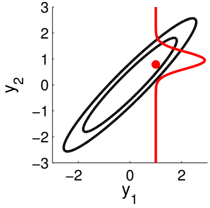
New visualisation



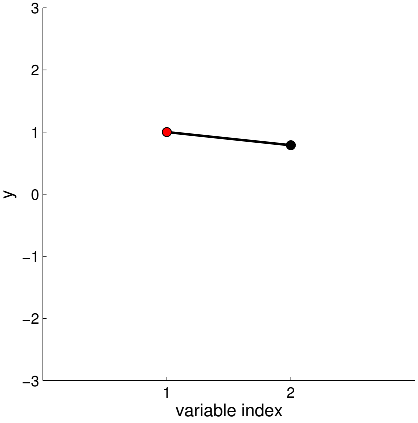
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



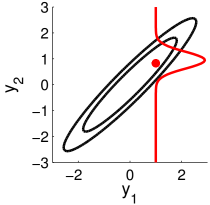
New visualisation



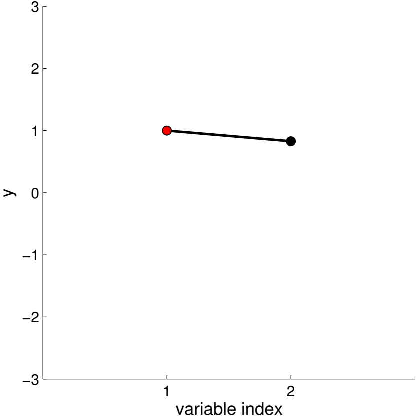
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



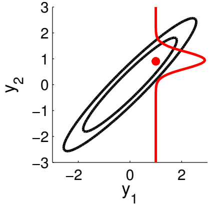
New visualisation



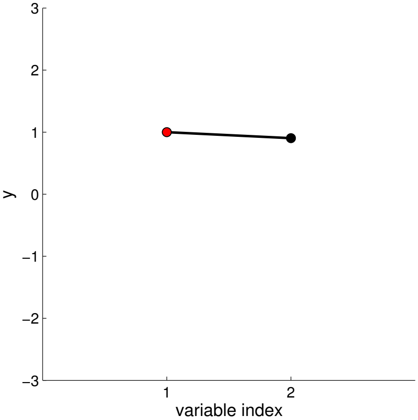
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



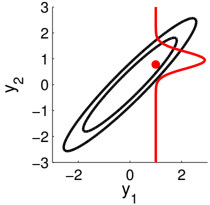
New visualisation



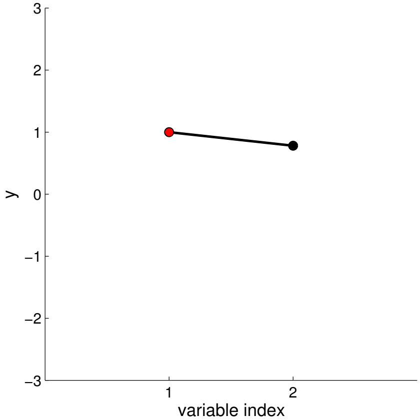
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



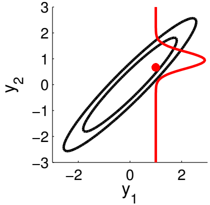
New visualisation



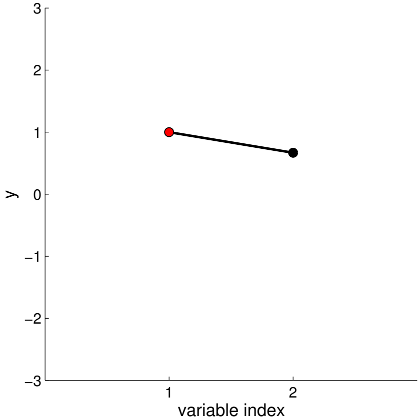
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



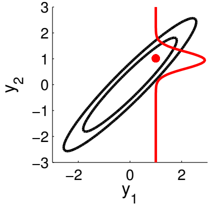
New visualisation

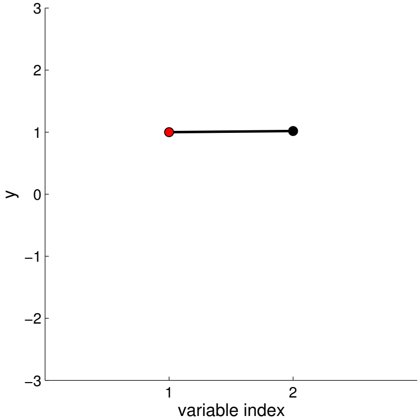


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



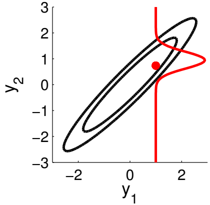
New visualisation



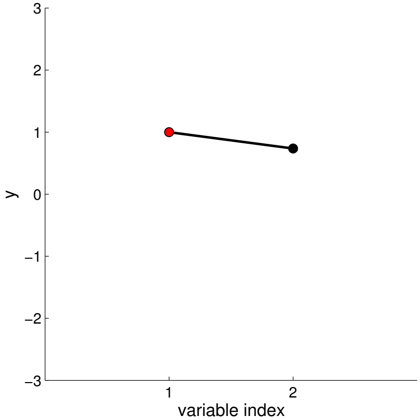


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

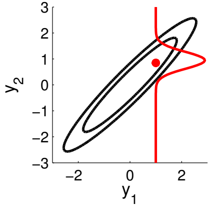
New visualisation



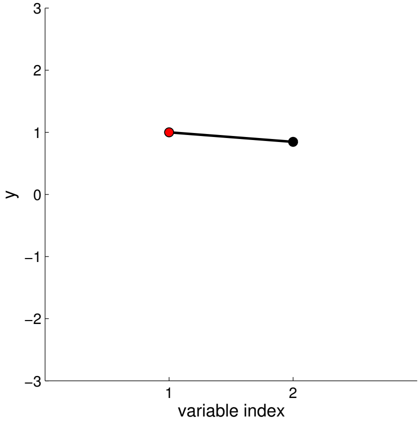
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



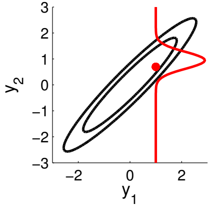
New visualisation



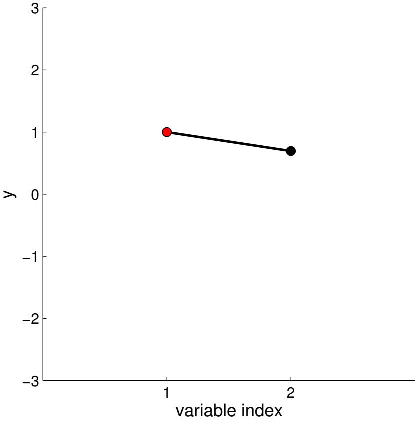
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



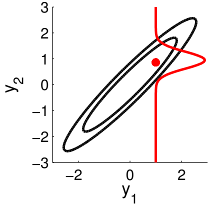
New visualisation



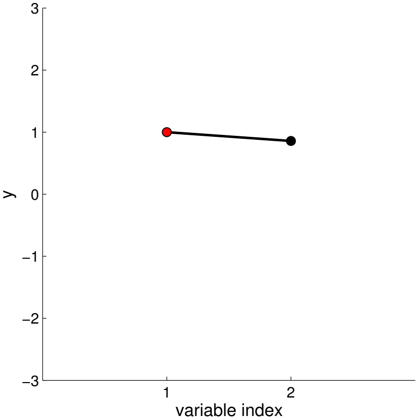
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



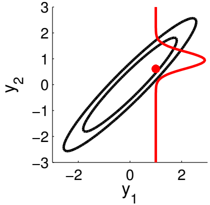
New visualisation



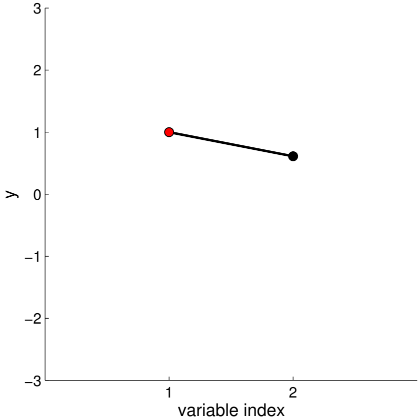
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



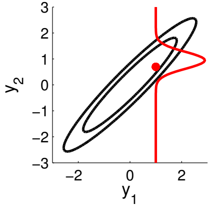
New visualisation



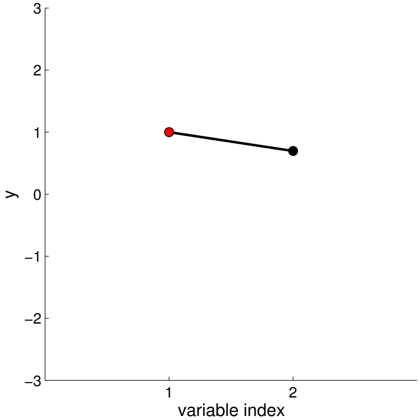
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



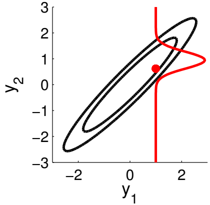
New visualisation

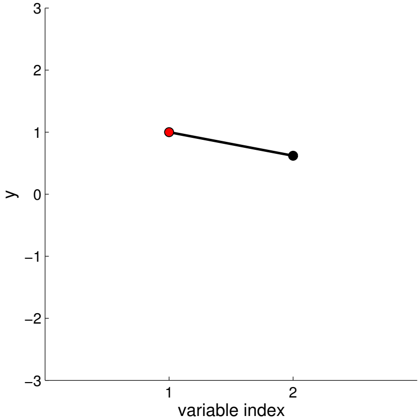


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



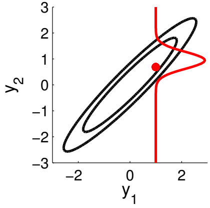
New visualisation



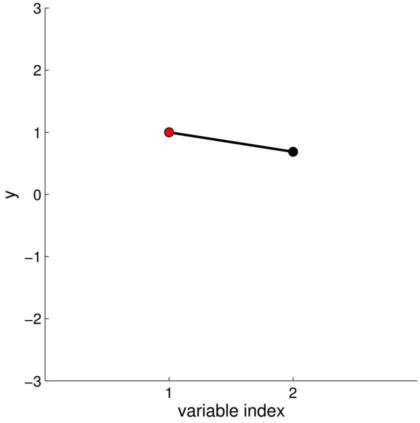


$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$

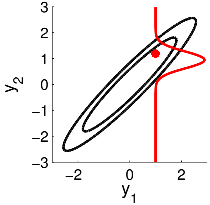
New visualisation



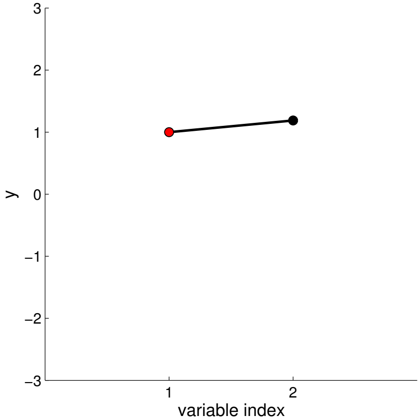
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



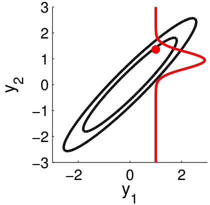
New visualisation



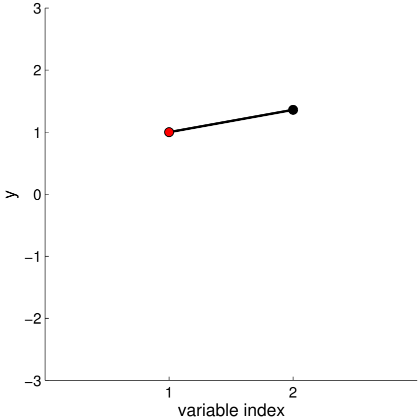
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



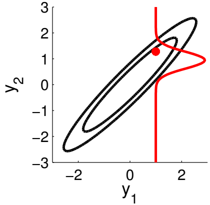
New visualisation



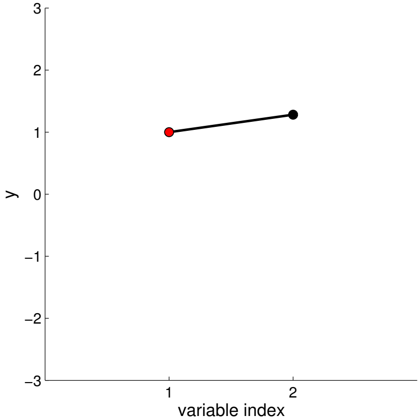
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



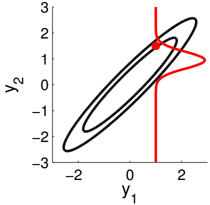
New visualisation



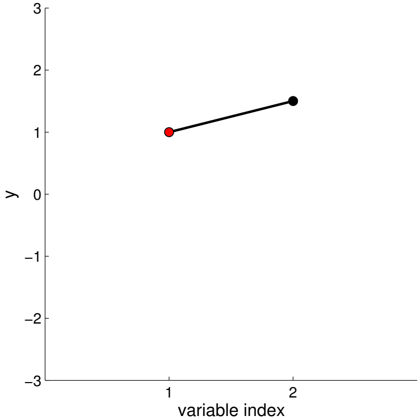
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



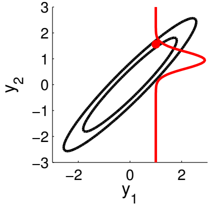
New visualisation



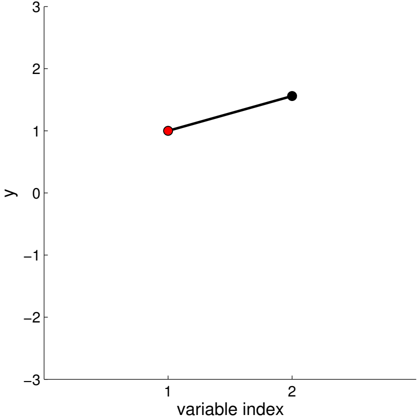
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



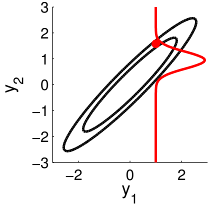
New visualisation



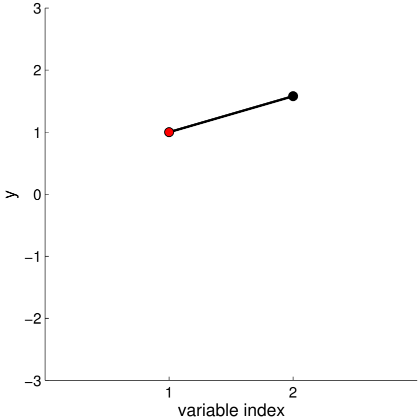
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



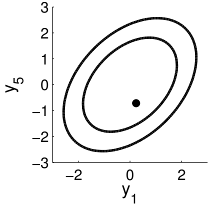
New visualisation



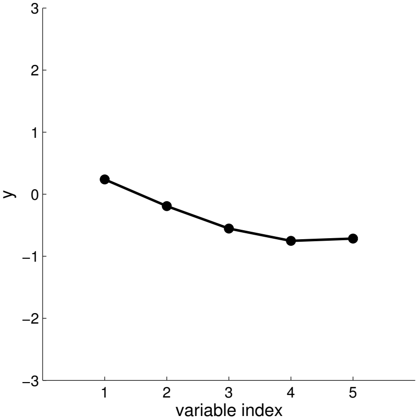
$$\Sigma = \begin{bmatrix} 1 & .9 \\ .9 & 1 \end{bmatrix}$$



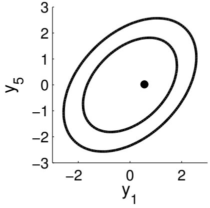
New visualisation

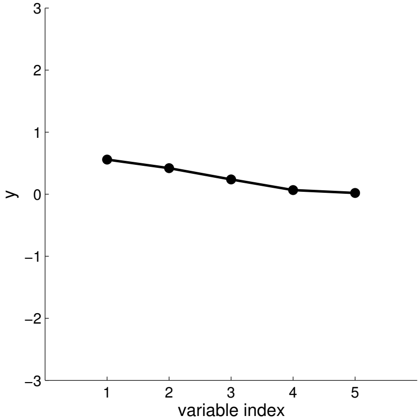


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



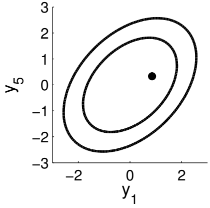
New visualisation

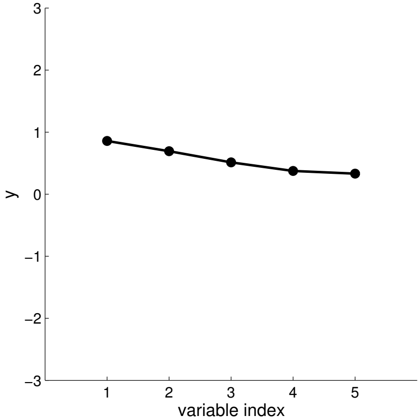




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

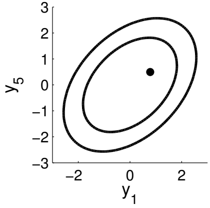
New visualisation



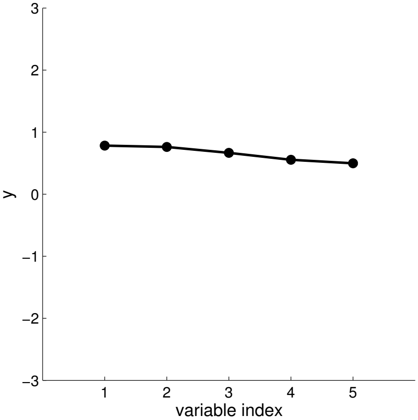


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

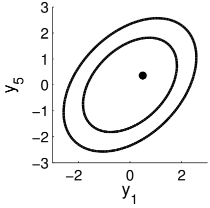
New visualisation

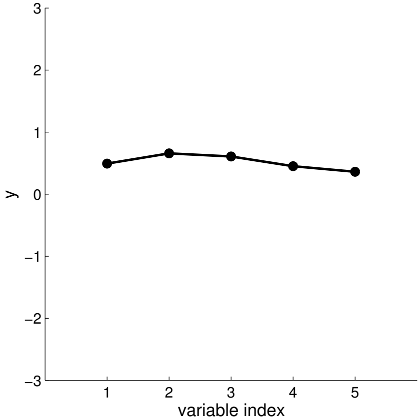


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



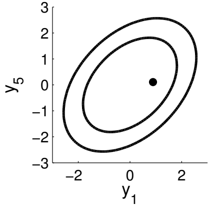
New visualisation



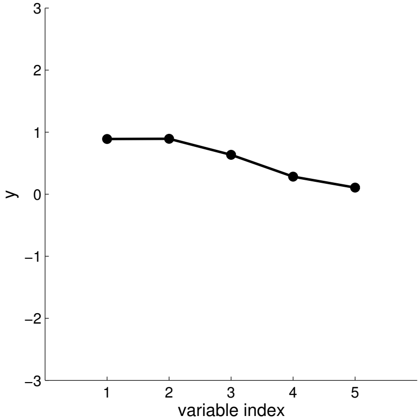


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

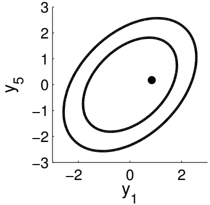
New visualisation

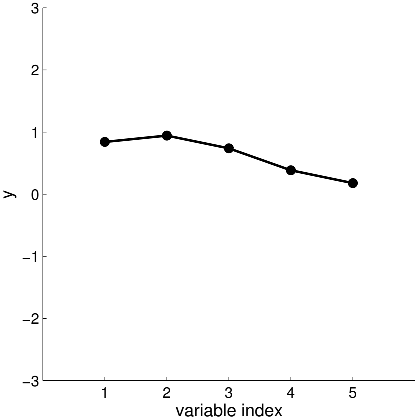


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



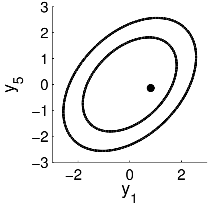
New visualisation

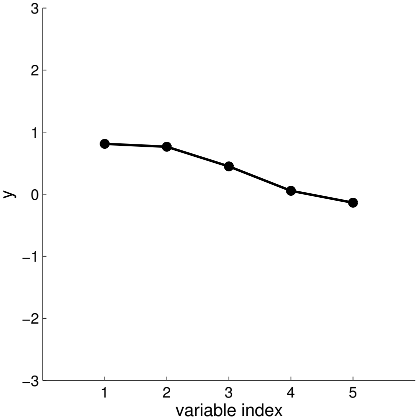




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

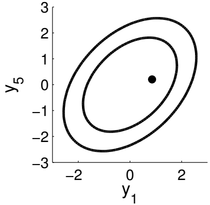
New visualisation

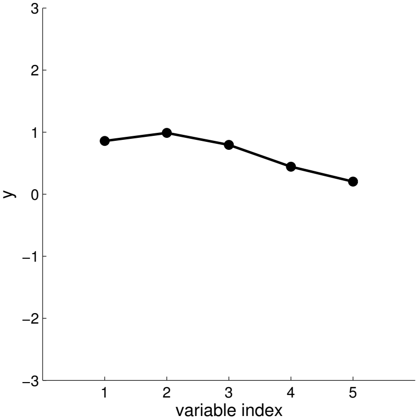




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

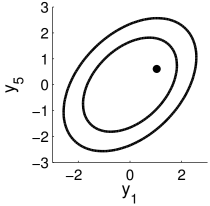
New visualisation

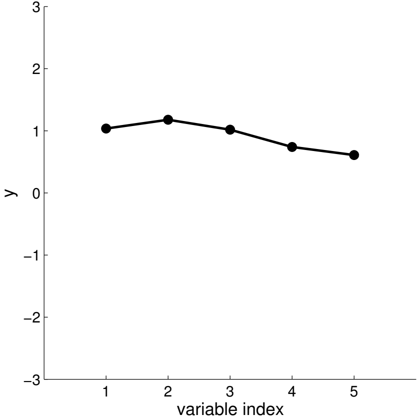




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

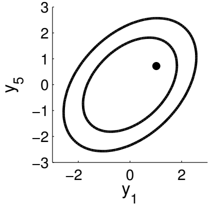
New visualisation

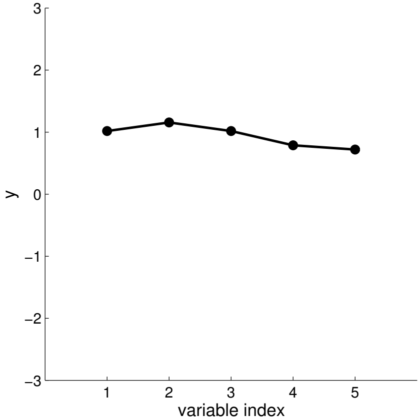




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

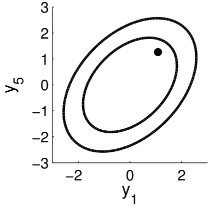
New visualisation

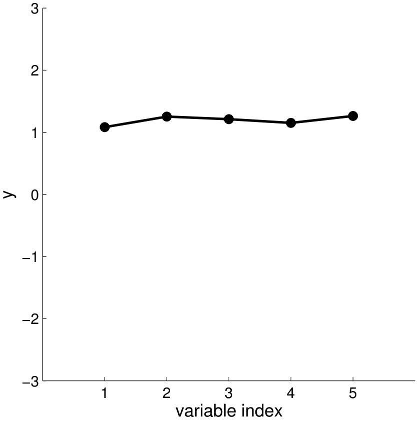




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

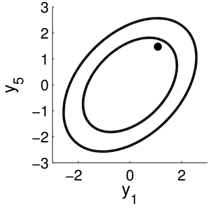
New visualisation

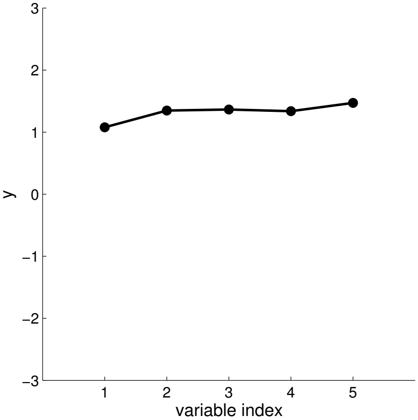




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

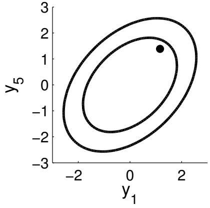
New visualisation

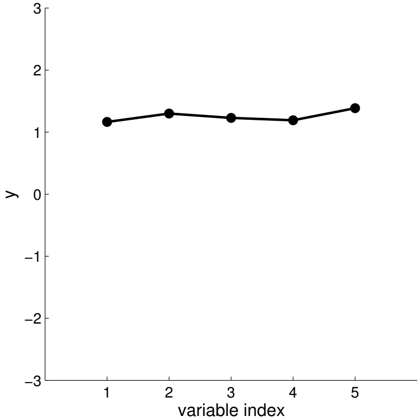




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

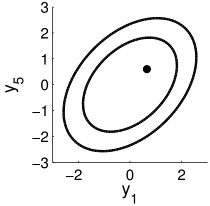
New visualisation

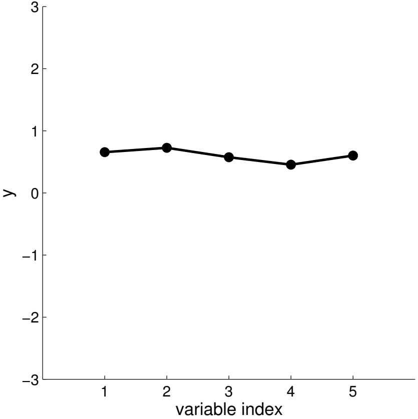




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

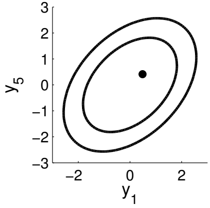
New visualisation



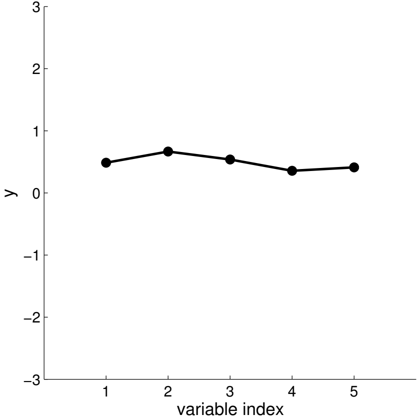


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

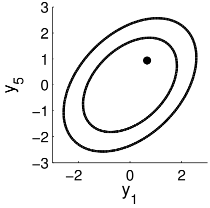
New visualisation

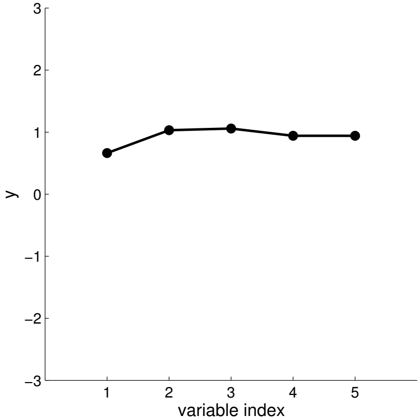


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



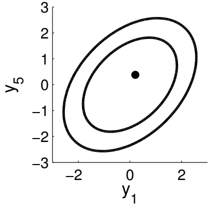
New visualisation

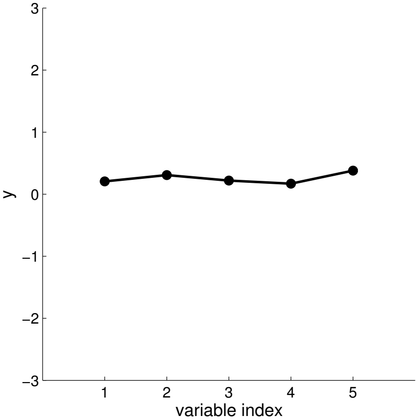




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

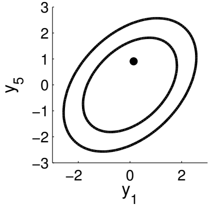
New visualisation



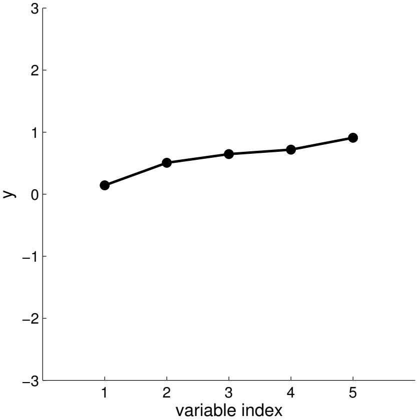


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

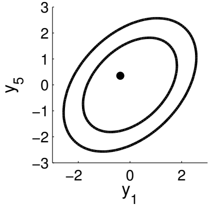
New visualisation

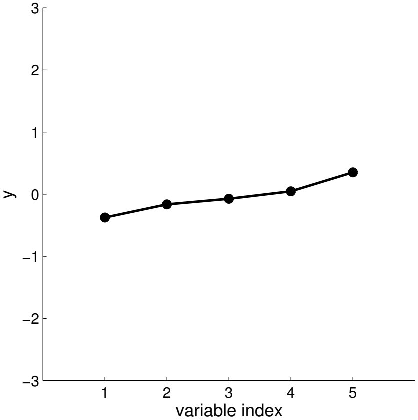


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



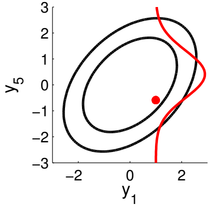
New visualisation

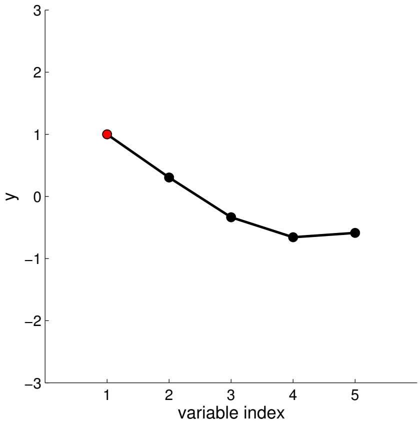




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

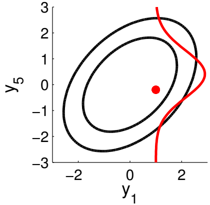
New visualisation

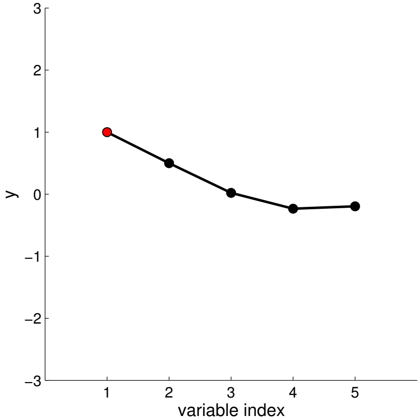




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

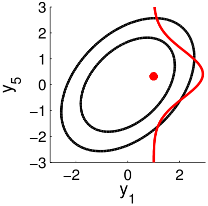
New visualisation

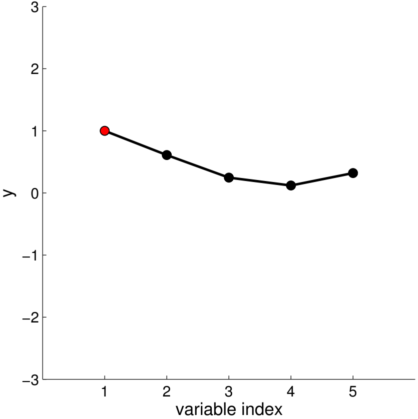




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

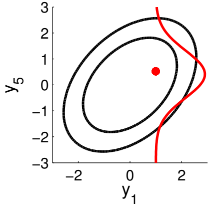
New visualisation

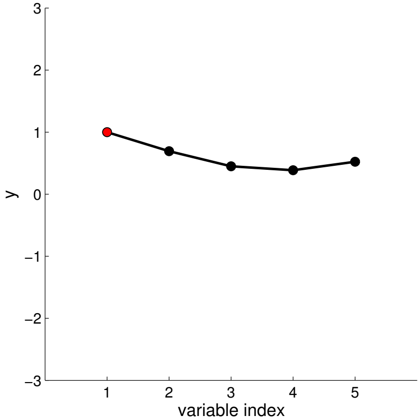




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

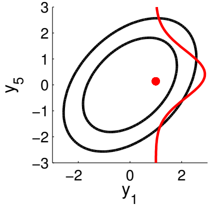
New visualisation

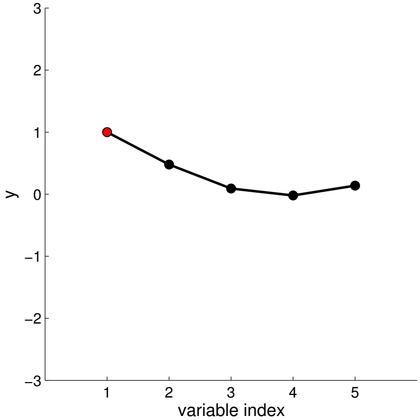




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

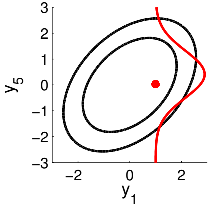
New visualisation



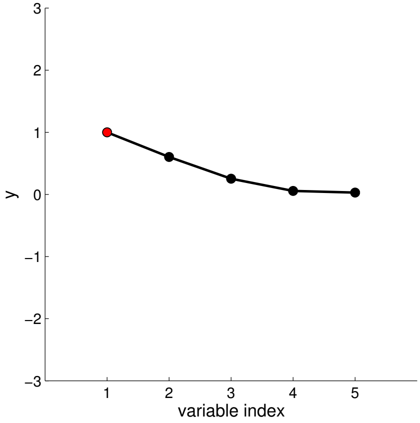


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

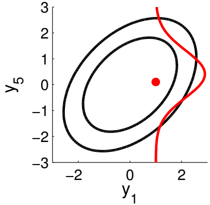
New visualisation



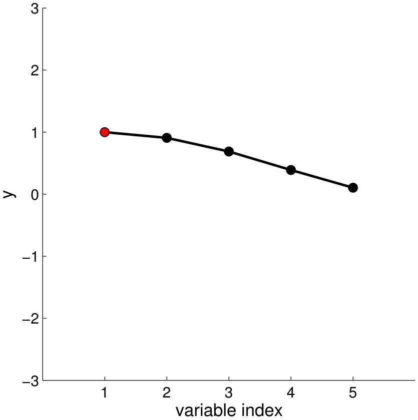
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



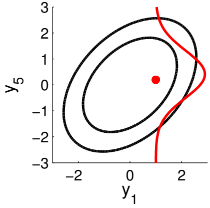
New visualisation

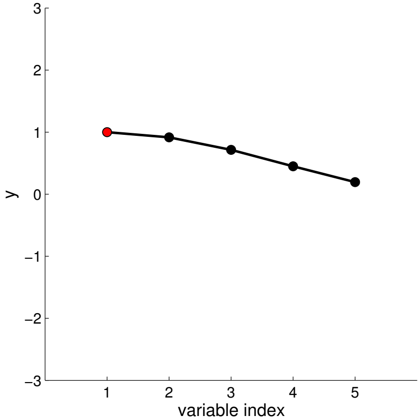


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



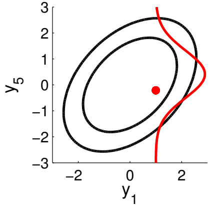
New visualisation

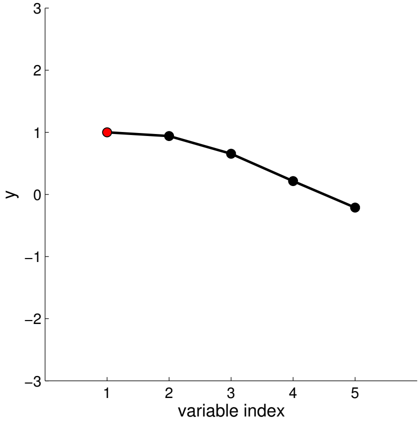




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

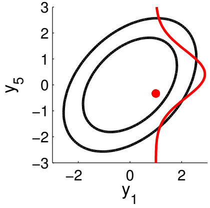
New visualisation

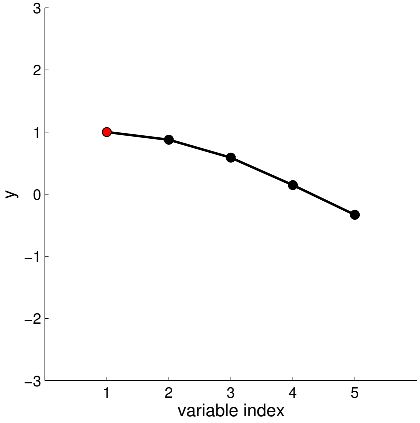




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

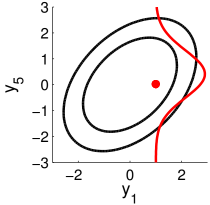
New visualisation

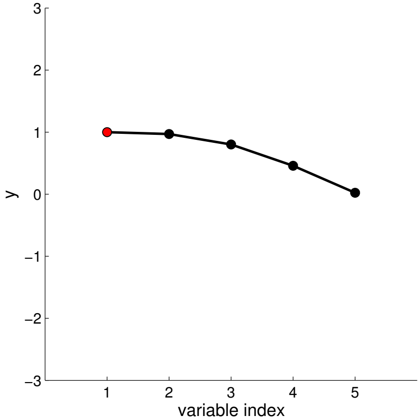




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

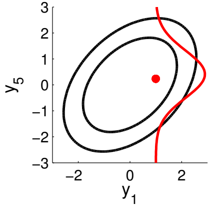
New visualisation

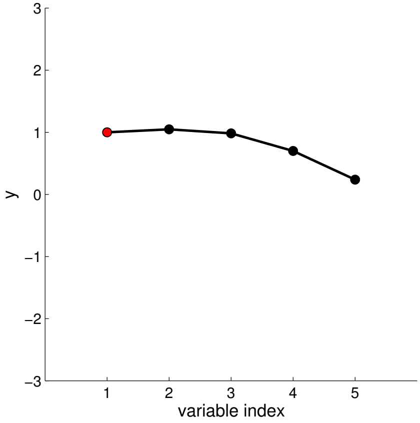




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

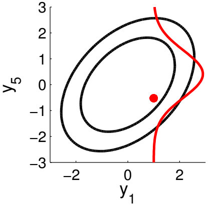
New visualisation

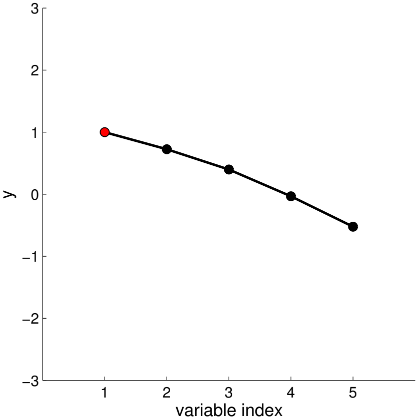




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

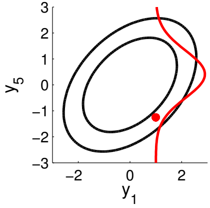
New visualisation

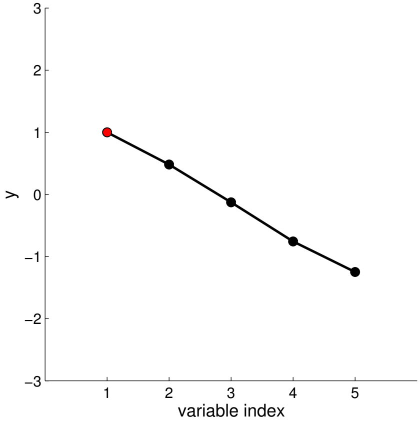




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

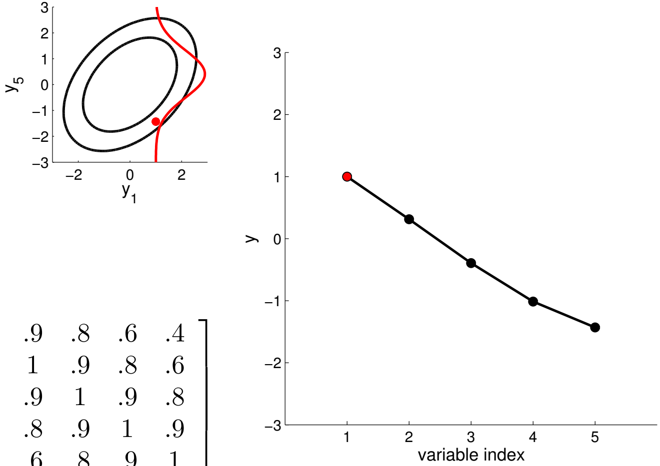
New visualisation





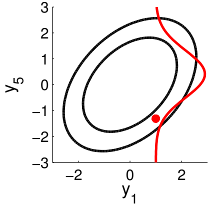
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

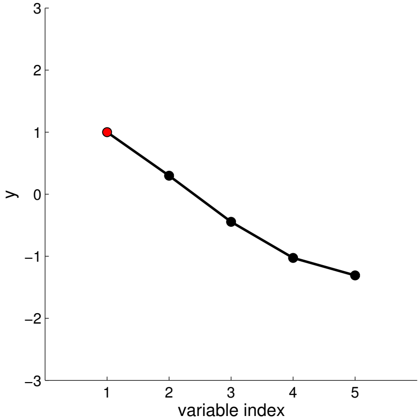
New visualisation



$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

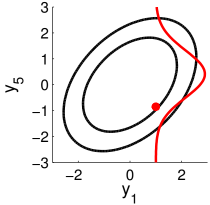
New visualisation

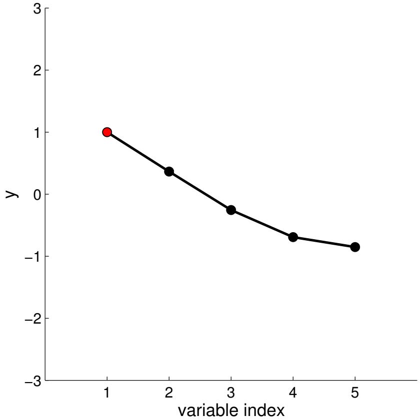




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

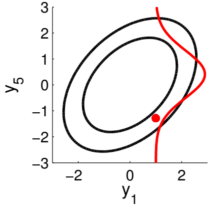
New visualisation

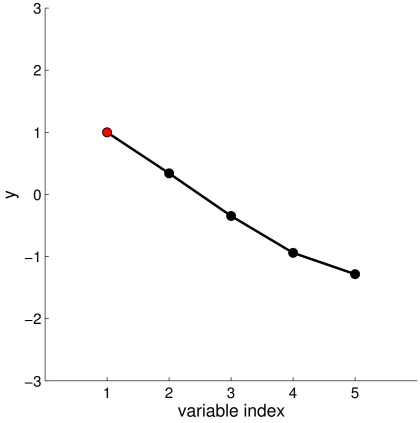




$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

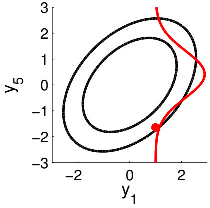
New visualisation



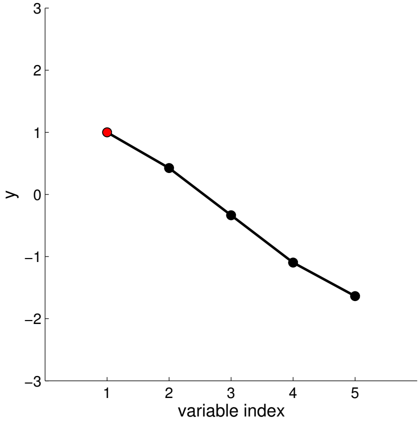


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

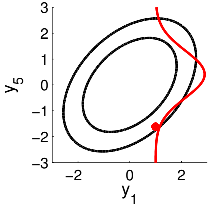
New visualisation



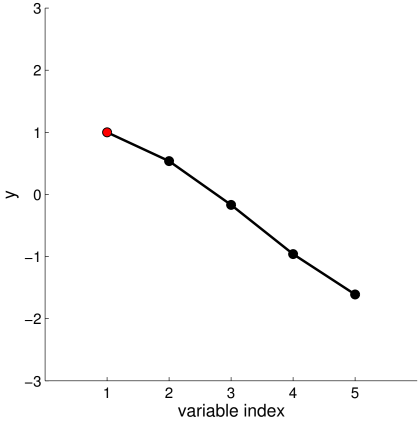
$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$



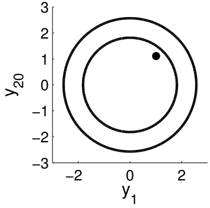
New visualisation

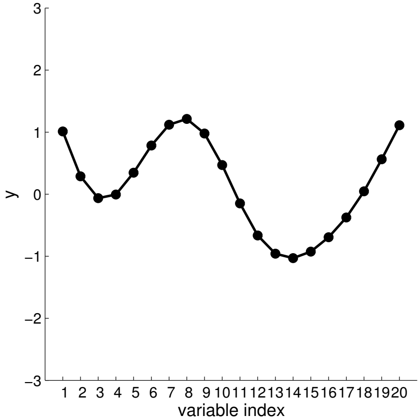


$$\Sigma = \begin{bmatrix} 1 & .9 & .8 & .6 & .4 \\ .9 & 1 & .9 & .8 & .6 \\ .8 & .9 & 1 & .9 & .8 \\ .6 & .8 & .9 & 1 & .9 \\ .4 & .6 & .8 & .9 & 1 \end{bmatrix}$$

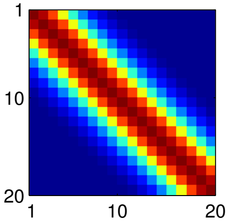


New visualisation

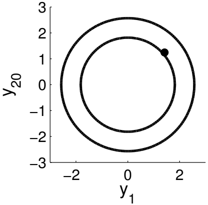




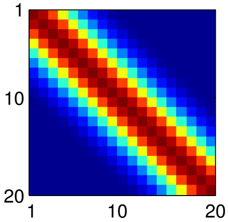
$\Sigma =$

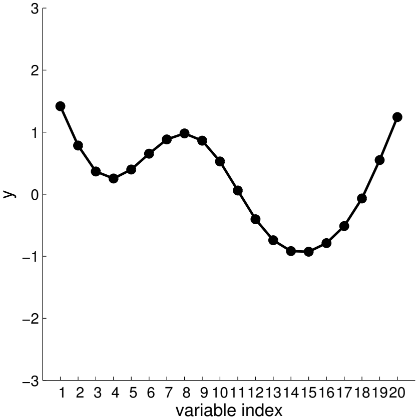


New visualisation

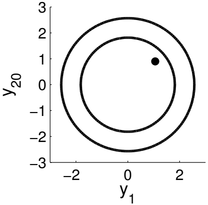


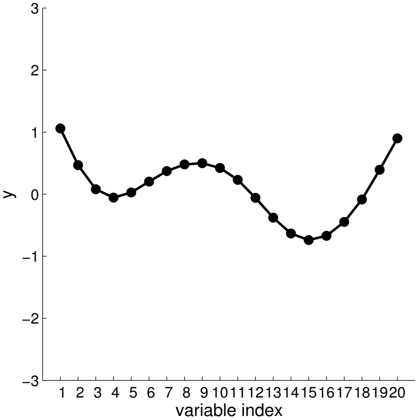
$\Sigma =$



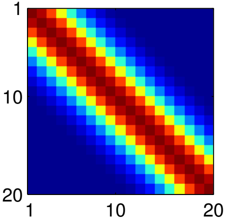


New visualisation

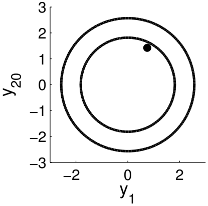


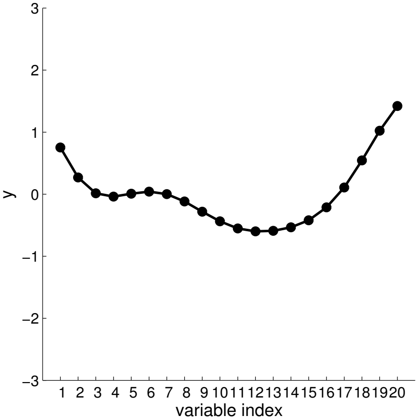


$\Sigma =$

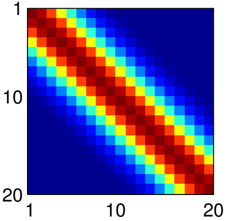


New visualisation

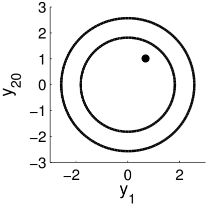


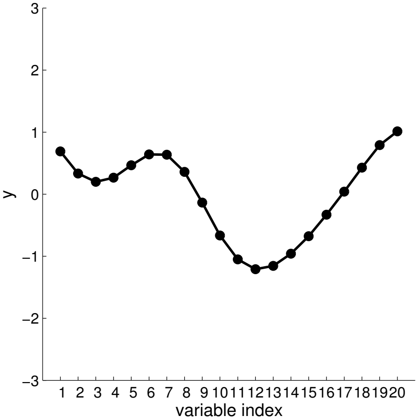


$\Sigma =$

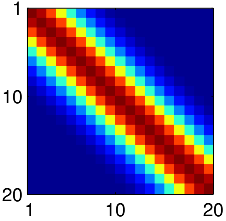


New visualisation

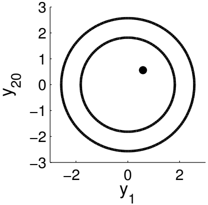




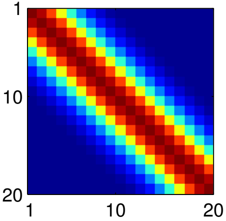
$\Sigma =$

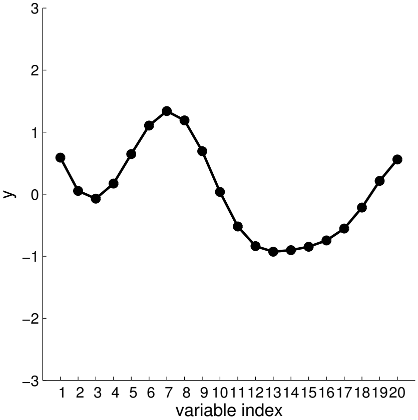


New visualisation

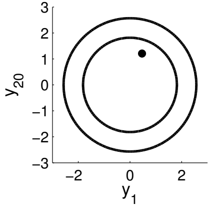


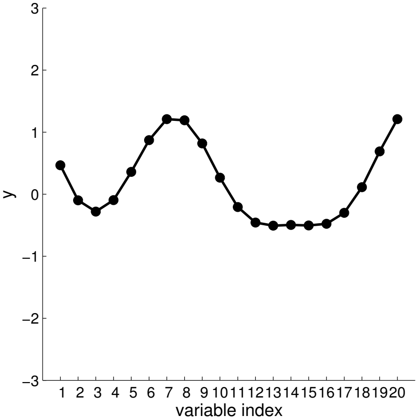
$$\Sigma =$$



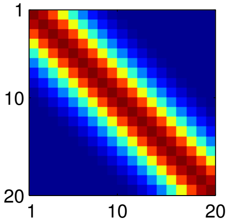


New visualisation

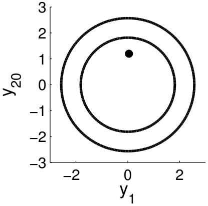




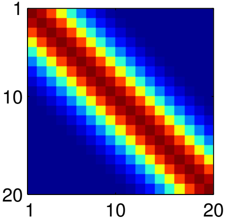
$\Sigma =$

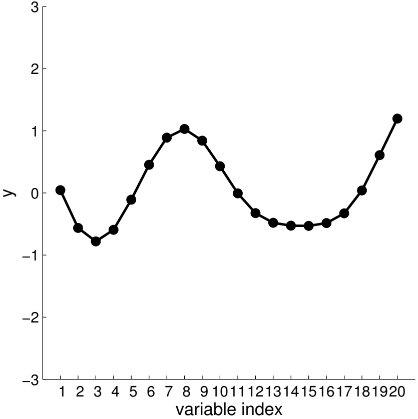


New visualisation

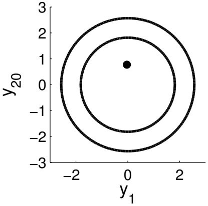


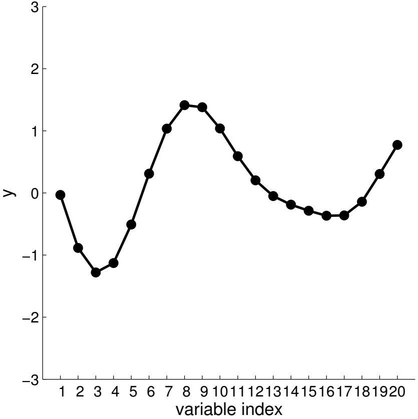
$$\Sigma =$$



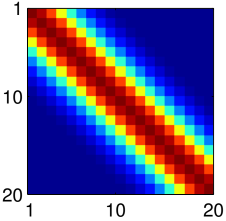


New visualisation

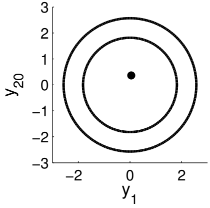


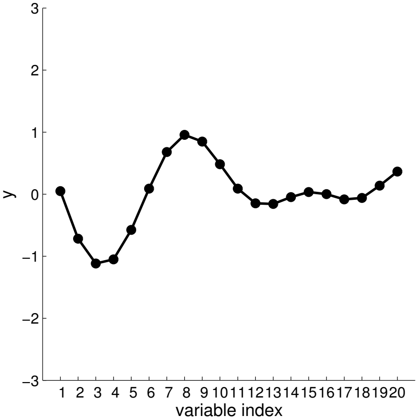


$\Sigma =$

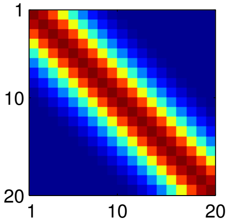


New visualisation

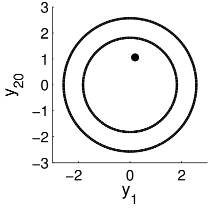




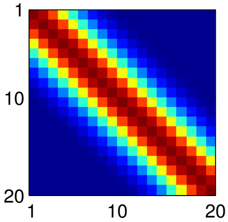
$\Sigma =$

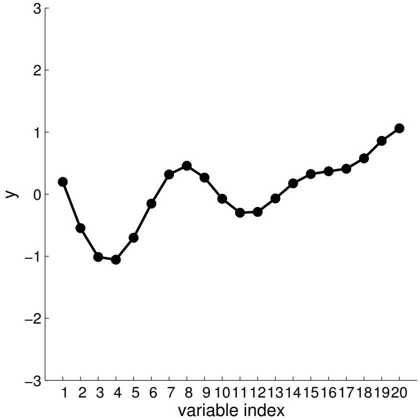


New visualisation

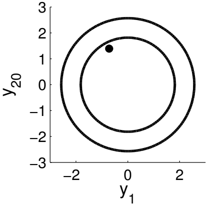


$\Sigma =$

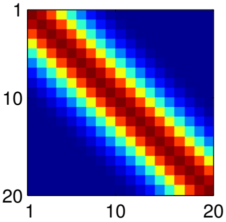


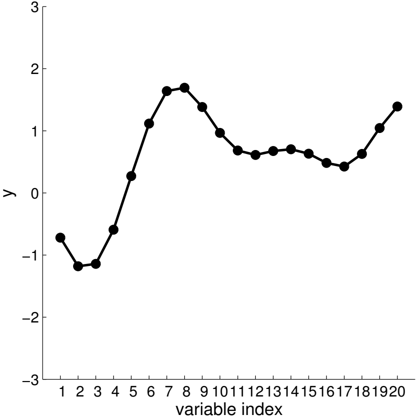


New visualisation

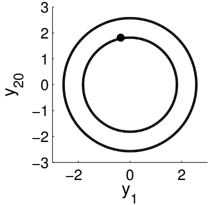


$\Sigma =$

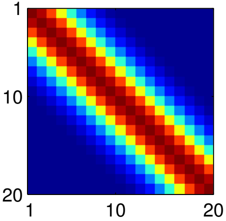


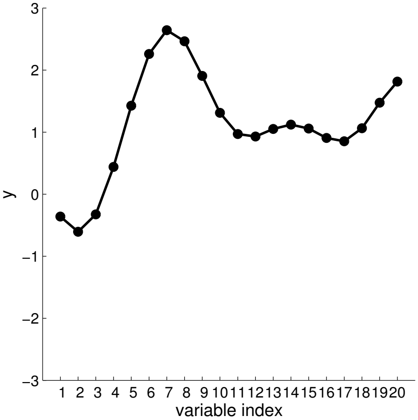


New visualisation

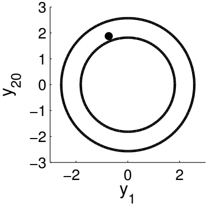


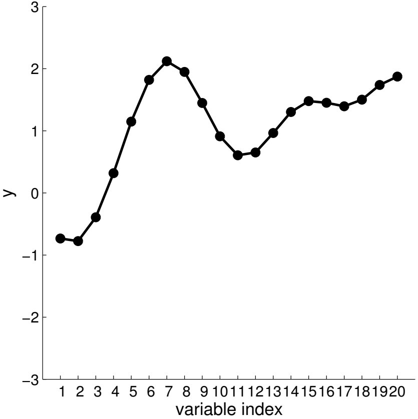
$$\Sigma =$$



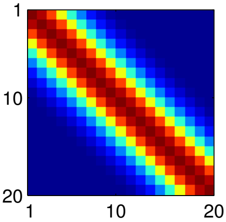


New visualisation

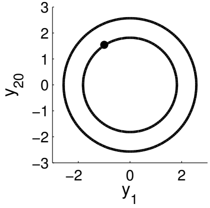




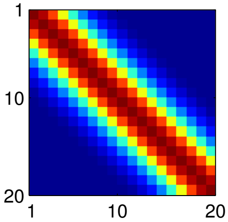
$$\Sigma =$$

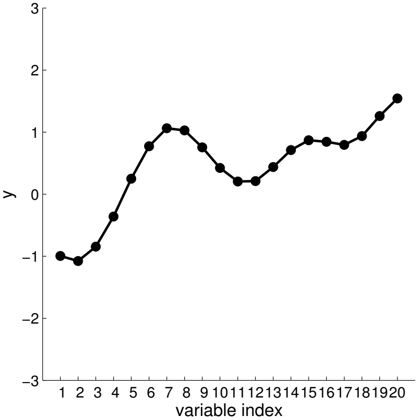


New visualisation

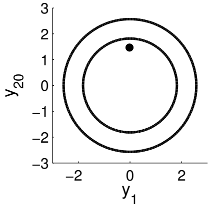


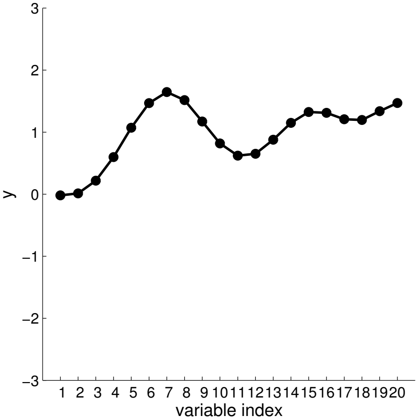
$$\Sigma =$$



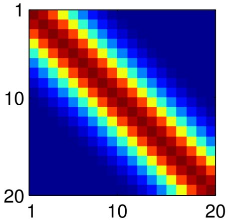


New visualisation

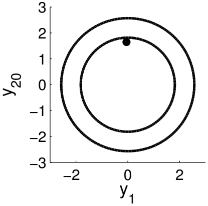


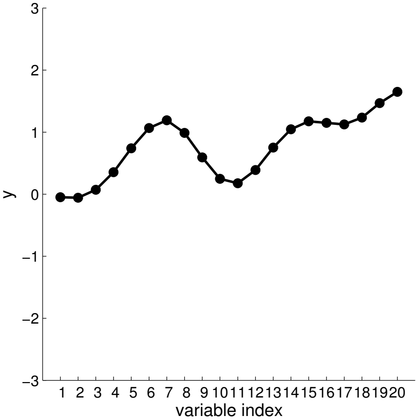


$\Sigma =$

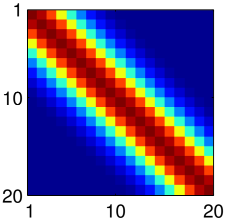


New visualisation

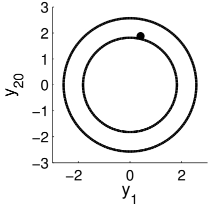


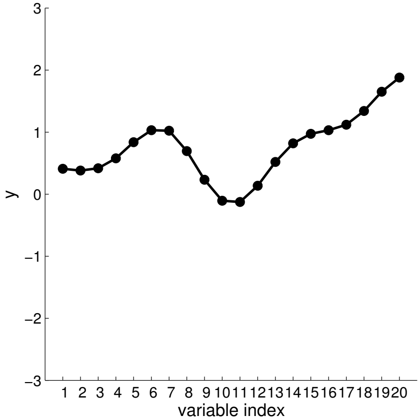


$$\Sigma =$$

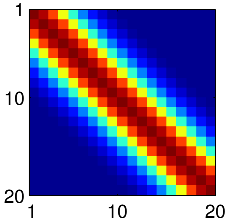


New visualisation

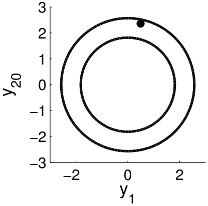




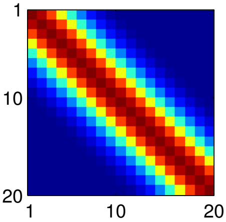
$\Sigma =$

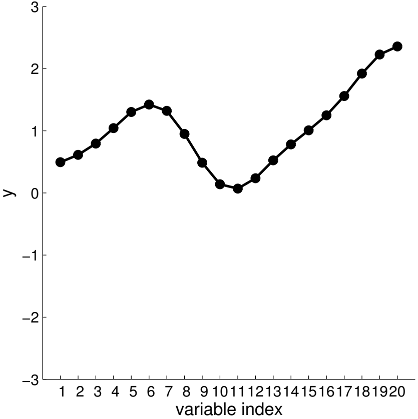


New visualisation

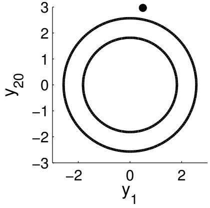


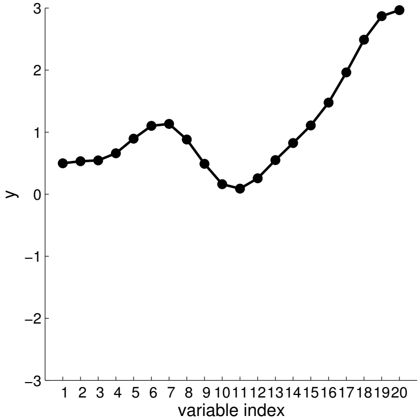
$\Sigma =$



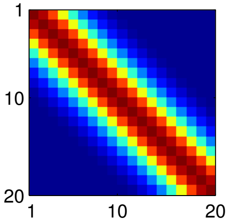


New visualisation

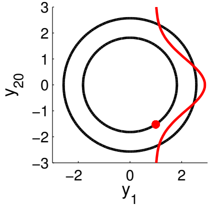




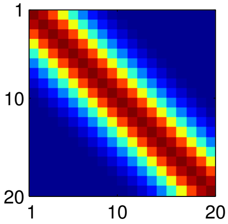
$\Sigma =$

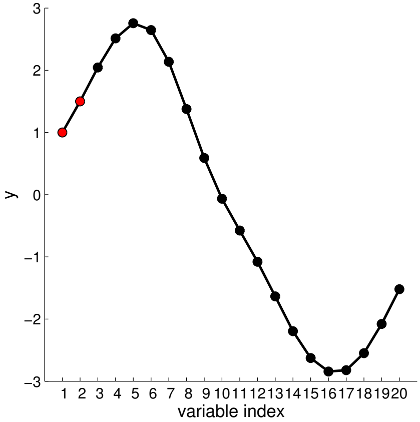


New visualisation

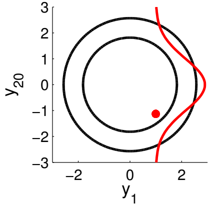


$$\Sigma =$$

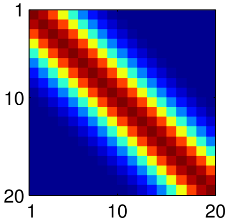


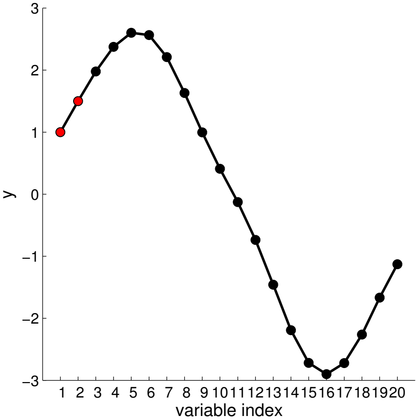


New visualisation

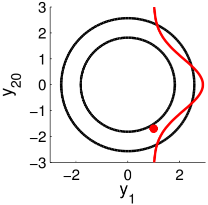


$\Sigma =$

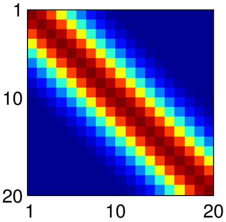


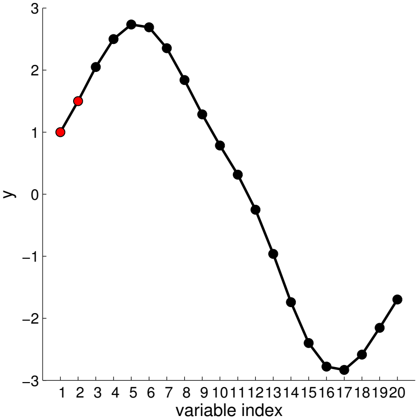


New visualisation

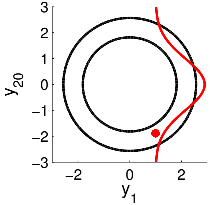


$\Sigma =$

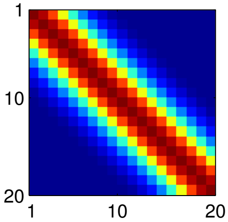


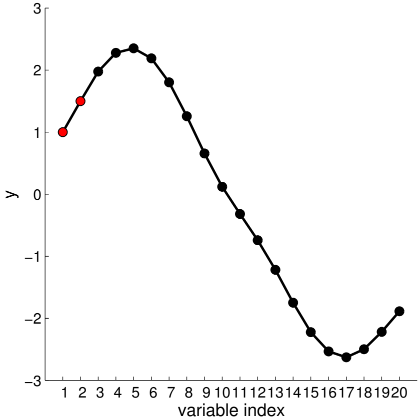


New visualisation

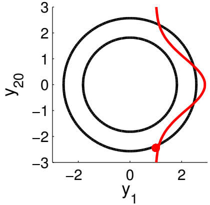


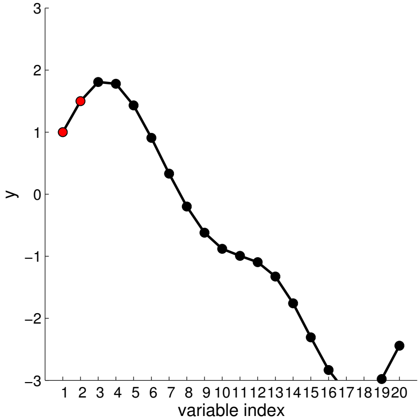
$\Sigma =$



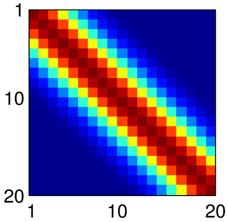


New visualisation

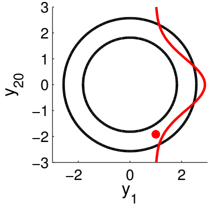


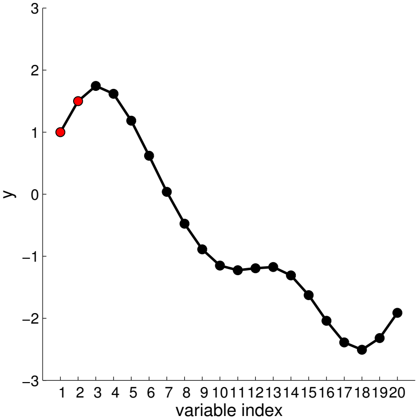


$\Sigma =$

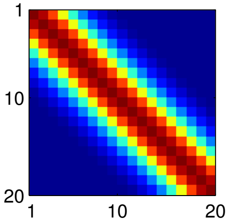


New visualisation

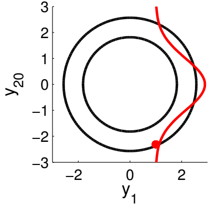




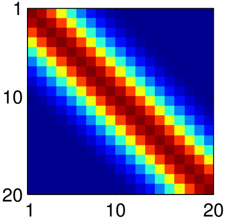
$\Sigma =$

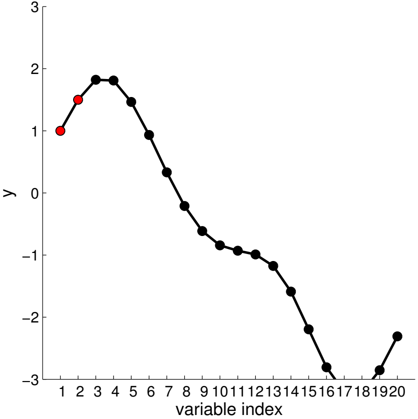


New visualisation

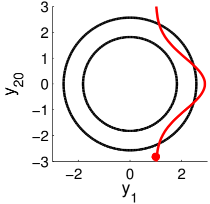


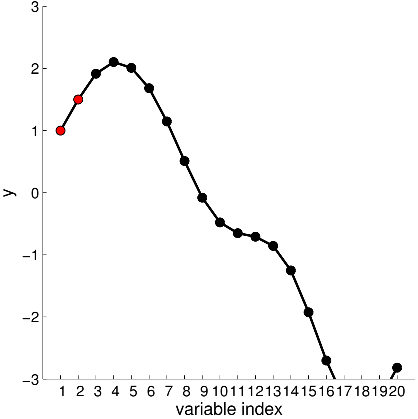
$\Sigma =$



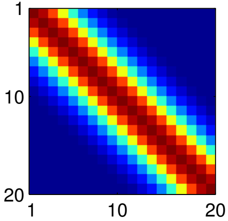


New visualisation

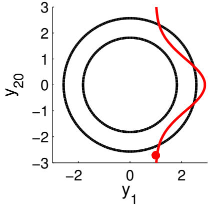




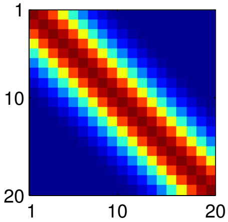
$$\Sigma =$$

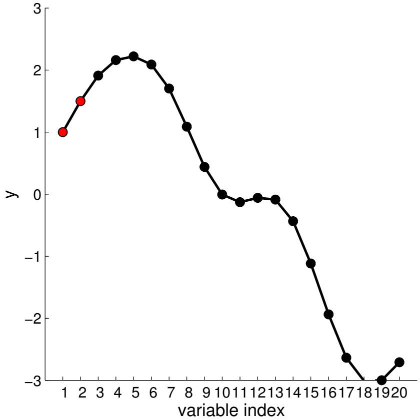


New visualisation

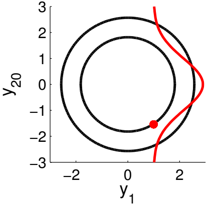


$$\Sigma =$$

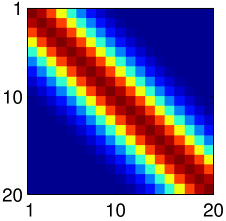


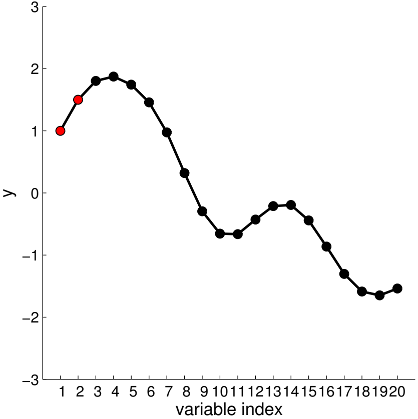


New visualisation

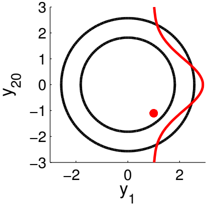


$\Sigma =$

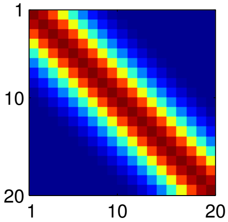


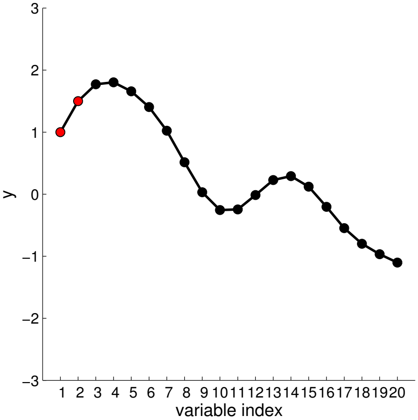


New visualisation

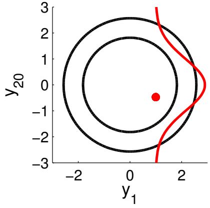


$\Sigma =$

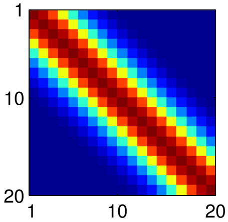


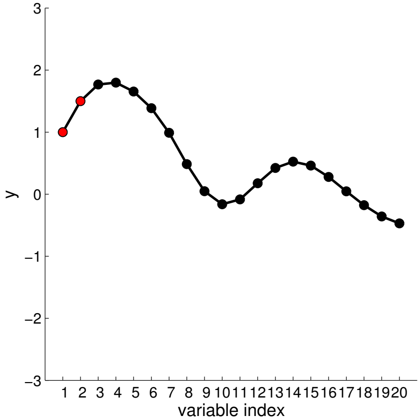


New visualisation

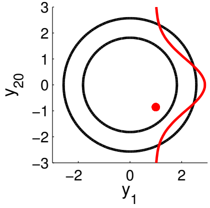


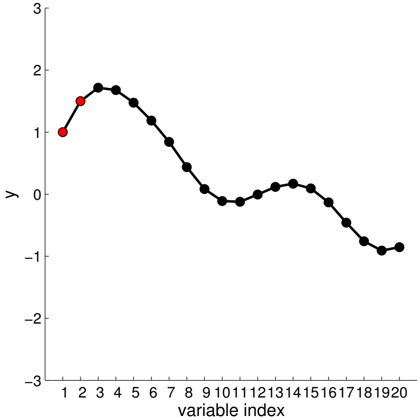
$\Sigma =$



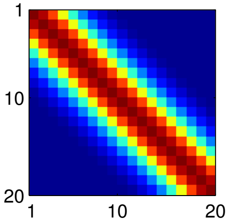


New visualisation

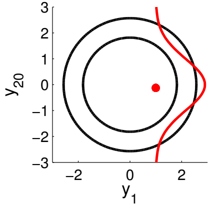


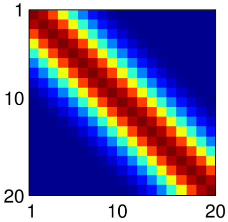


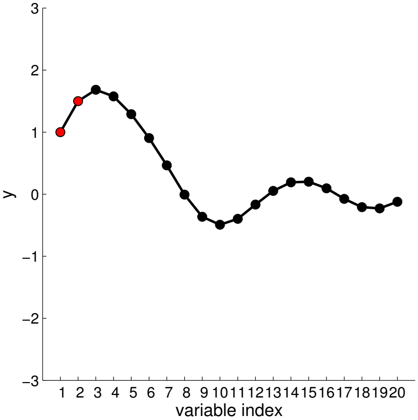
$\Sigma =$



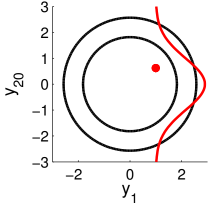
New visualisation

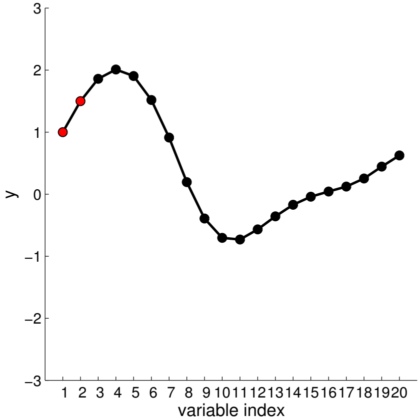


$\Sigma =$ 

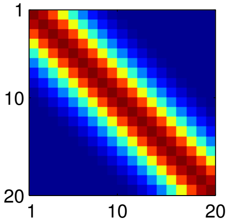


New visualisation

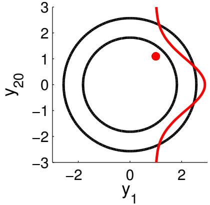


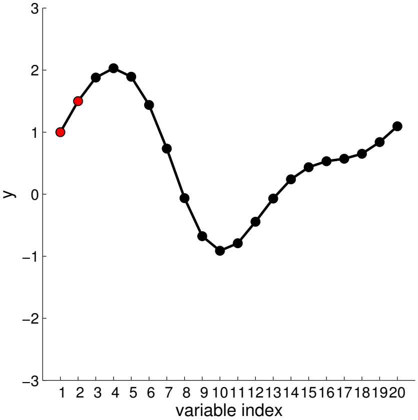


$$\Sigma =$$

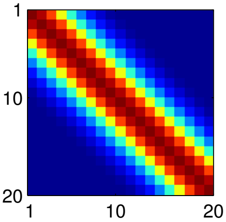


New visualisation

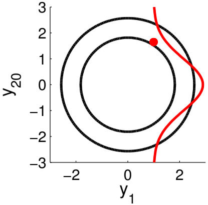


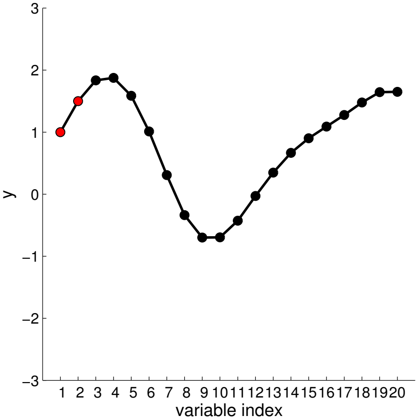


$$\Sigma =$$

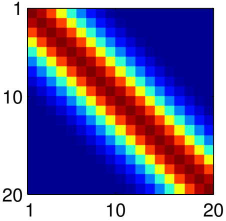


New visualisation



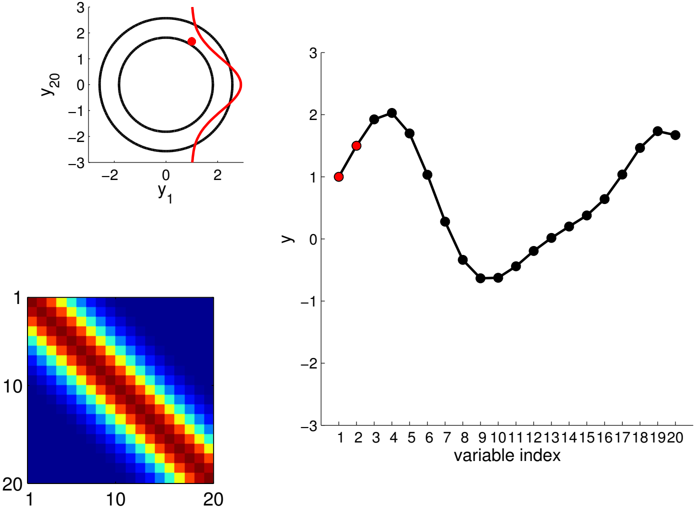


$\Sigma =$

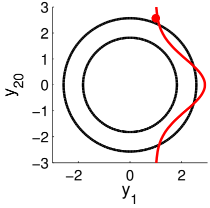


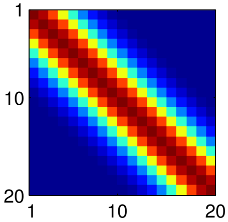
New visualisation

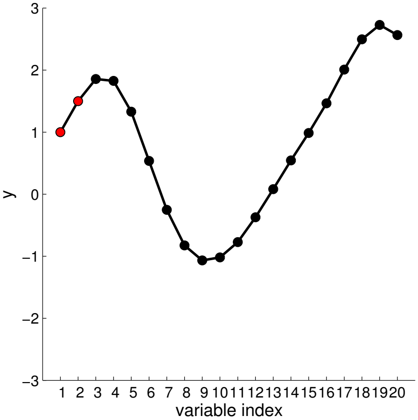
$\Sigma =$



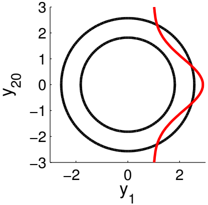
New visualisation



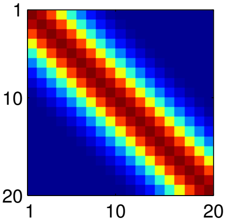
$\Sigma =$ 

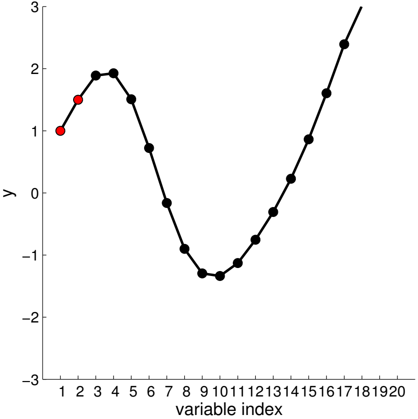


New visualisation

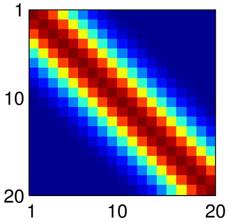


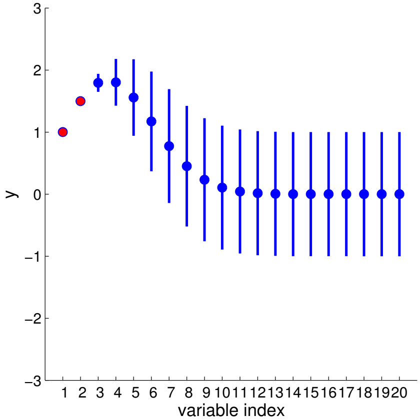
$$\Sigma =$$





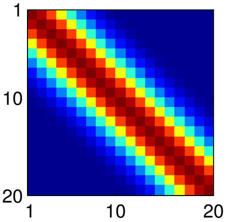
Regression using Gaussians

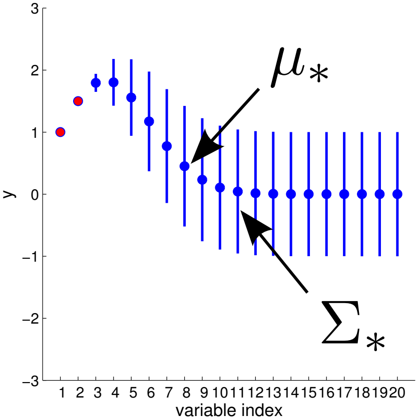
$\Sigma =$ 



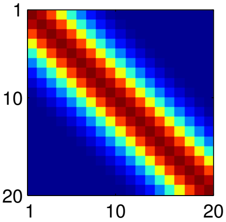
Regression using Gaussians

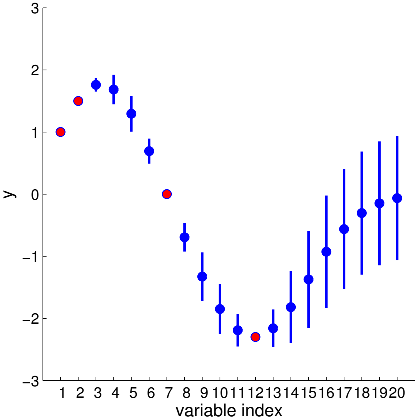
$\Sigma =$





Regression using Gaussians

$\Sigma =$ 

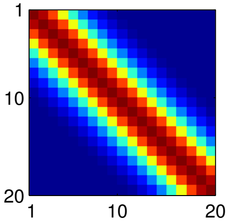


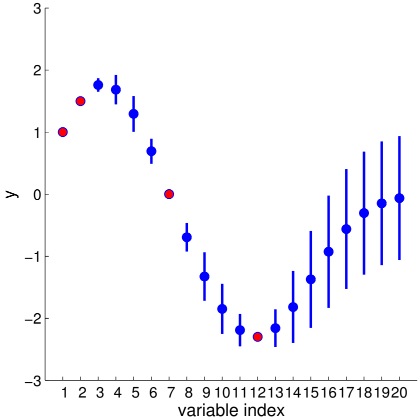
Regression using Gaussians

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

$\Sigma =$



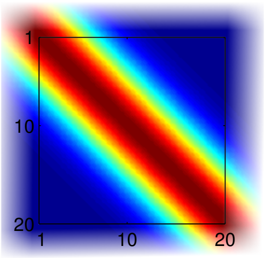


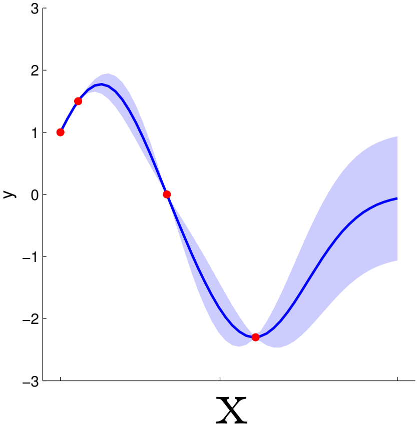
Regression: probabilistic inference in function space

$$\Sigma(x_1, x_2) = K(x_1, x_2) + I\sigma_y^2$$

$$K(x_1, x_2) = \sigma^2 e^{-\frac{1}{2l^2}(x_1 - x_2)^2}$$

$\Sigma =$





Regression: probabilistic inference in function space

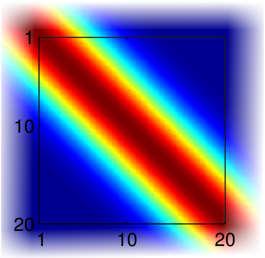
Non-parametric (∞-parametric)

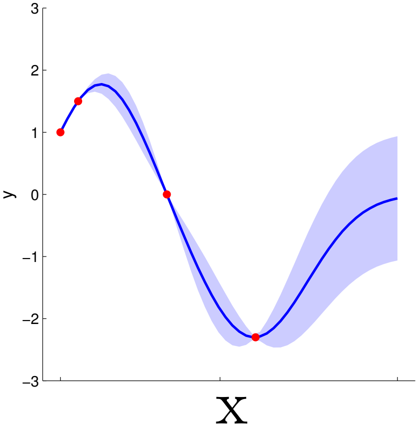
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$





Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

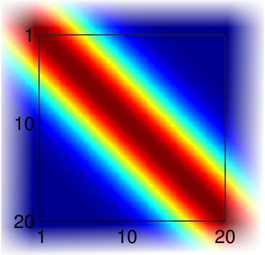
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

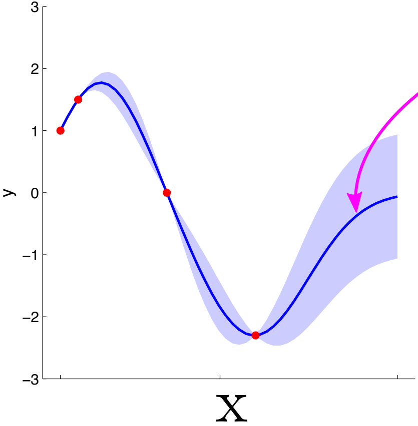
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



function estimate
with uncertainty



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

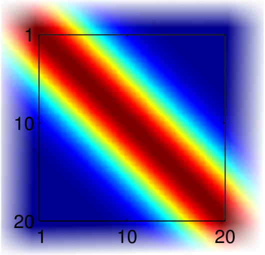
Non-parametric (∞-parametric)

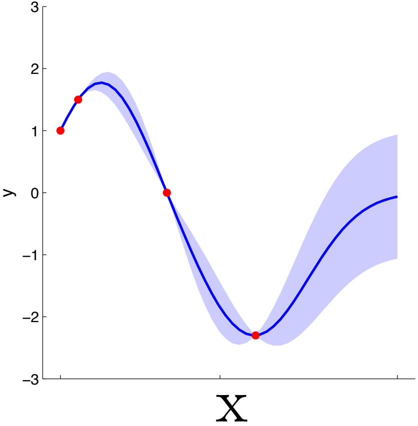
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

observation noise

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

Non-parametric (∞-parametric)

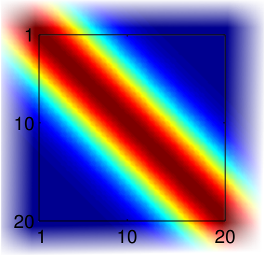
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

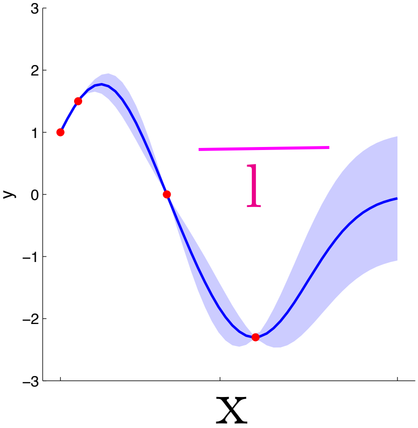
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

horizontal-scale

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Regression: probabilistic inference in function space

Non-parametric (∞-parametric)

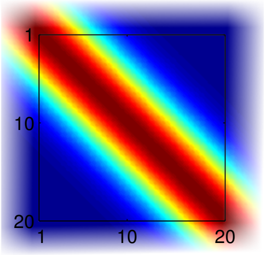
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

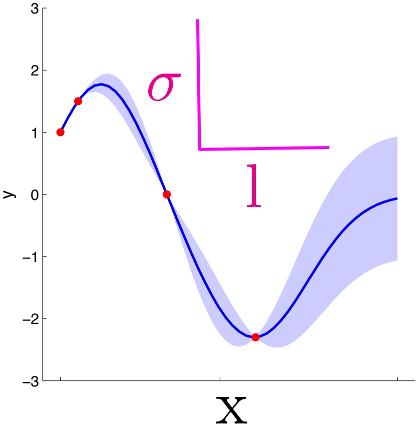
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

vertical-scale   horizontal-scale

$\Sigma =$





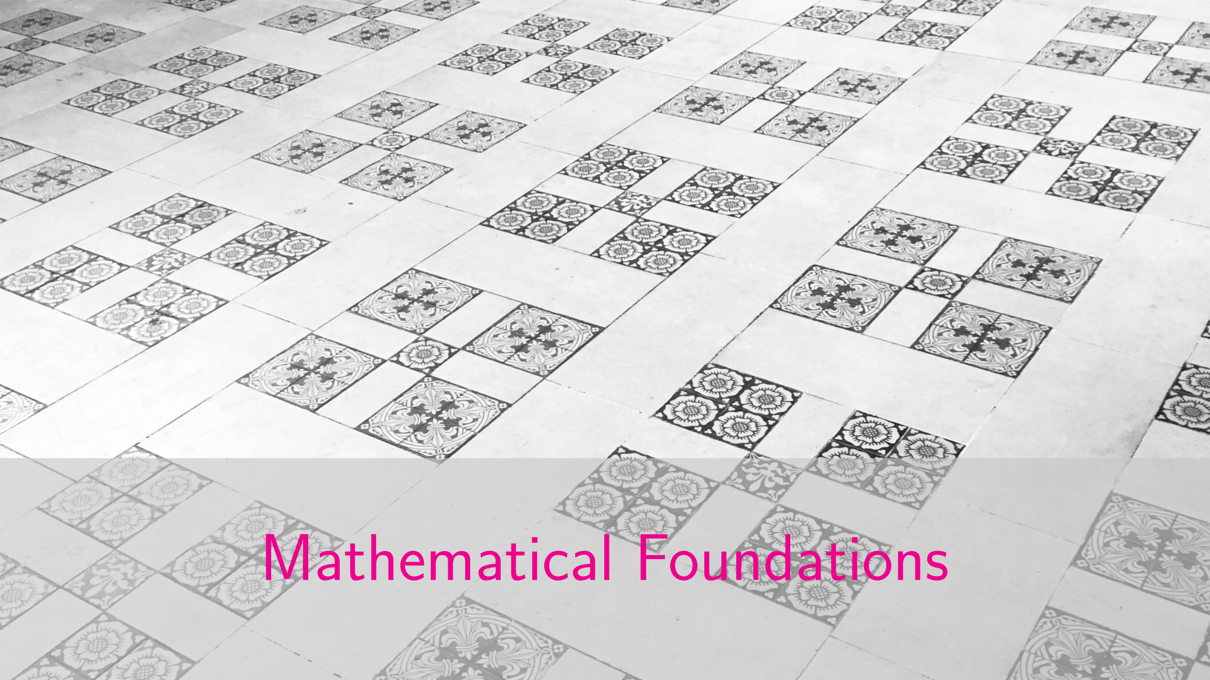
Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

Outline of the tutorial

- **An Introduction to GPs**
  - Mathematical foundations
  - Hyper-parameter learning
  - Covariance functions
  - Multi-dimensional inputs
- **Using GPs: Models, Applications and Connections**
  - Models and more on covariance functions
  - Applications
  - Connections
- **GPs for large data and non-linear models**
  - Scaling through pseudo-data
  - Variational Inference
  - General Approximate inference



Mathematical Foundations

Mathematical Foundations: Definition

Gaussian process = generalisation of multivariate Gaussian distribution to infinitely many variables.

Definition: a Gaussian process is a collection of random variables, any finite number of which have (consistent) Gaussian distributions.

A Gaussian distribution is fully specified by a mean vector, $\boldsymbol{\mu}$, and covariance matrix Σ:

$$\mathbf{f} = (\mathsf{f}_1, \ldots, \mathsf{f}_n) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \quad \text{indices} \quad i = 1, \ldots, n$$

A Gaussian process is fully specified by a mean function $m(\mathbf{x})$ and covariance function $\mathsf{K}(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), \mathsf{K}(\mathbf{x}, \mathbf{x}')\right), \quad \text{indices} \quad \mathbf{x}$$

Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

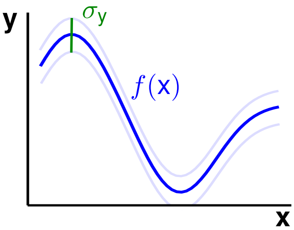
Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)

$$y(x) = f(x) + \epsilon \sigma_y$$

$$p(\epsilon) = \mathcal{N}(\epsilon; 0, 1)$$



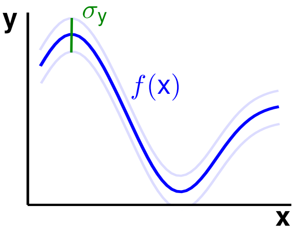
Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}) + \epsilon\sigma_{\mathrm{y}}$$

$$p(\epsilon) = \mathcal{N}(\epsilon; 0, 1)$$



place GP prior over the non-linear function

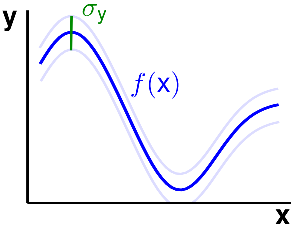
$$p(f(\mathrm{x})|\theta) = \mathcal{GP}(f(\mathrm{x}); 0, \mathrm{K}_{\theta}(\mathrm{x}, \mathrm{x}'))$$

$$\mathrm{K}(\mathrm{x}, \mathrm{x}') = \sigma^2 \exp\left(-\frac{1}{2l^2}(\mathrm{x} - \mathrm{x}')^2\right)$$ (smoothly wiggling functions expected)

Mathematical Foundations: Regression

Q1. What's the formal justification for how we were using GPs for regression?

generative model (like non-linear regression)



$$y(x) = f(x) + \epsilon\sigma_y$$

$$p(\epsilon) = \mathcal{N}(\epsilon; 0, 1)$$

place GP prior over the non-linear function

$$p(f(x)|\theta) = \mathcal{GP}(f(x); 0, K_\theta(x, x'))$$

$$K(x, x') = \sigma^2 \exp\left(-\frac{1}{2l^2}(x - x')^2\right)$$ (smoothly wiggling functions expected)

sum of Gaussian variables = Gaussian: induces a GP over $y(x)$

$$p(y(x)|\theta) = \mathcal{GP}(y(x); 0, K_\theta(x, x') + I\sigma_y^2)$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) \mathrm{d}\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} , \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix} \right)$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) \mathrm{d}\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} , \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix} \right)$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A)$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A)$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\mathsf{T} & \mathsf{C} \end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, \mathsf{A})$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A)$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1}B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1} B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

not the same

$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}^{-1} = \begin{bmatrix} (A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1} B)^{-1} & (C - BA^{-1}B^\top)^{-1} \end{bmatrix}$$

Mathematical Foundations: Marginalisation

Q2. A GP is "like" a Gaussian distribution with an infinitely long mean vector and an "infinite by infinite" covariance matrix, so how do we represent it on a computer?

We are saved by the marginalisation property:

$$p(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2$$

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2\end{bmatrix}; \begin{bmatrix}\mathbf{a} \\ \mathbf{b}\end{bmatrix}, \begin{bmatrix}A & B \\ B^\top & C\end{bmatrix}\right) \implies p(\mathbf{y}_1) = \mathcal{N}(\mathbf{y}_1; \mathbf{a}, A) \implies \mathcal{P} = A^{-1}$$

$\implies$ Only need to represent finite dimensional projections of GPs on computer.

Q3. Why do we use covariances and not precisions to express modelling choices?

$$\mathcal{P} = \Sigma^{-1}$$

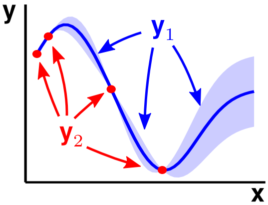
not the same

$$\begin{bmatrix}A & B \\ B^\top & C\end{bmatrix}^{-1} = \begin{bmatrix}(A - B^\top C^{-1} B)^{-1} & -A^{-1}B(C - BA^{-1}B^\top)^{-1} \\ -C^{-1}B(A - B^\top C^{-1}B)^{-1} & (C - BA^{-1}B^\top)^{-1}\end{bmatrix}$$

$\implies$ Entries in a precision matrix depend on what other data we are considering

Mathematical Foundations: Prediction

Q4. How do we make predictions?

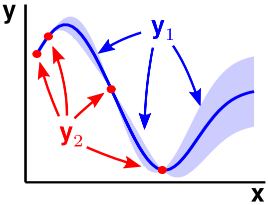


;

Mathematical Foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^{\top} & C \end{bmatrix} \right)$$



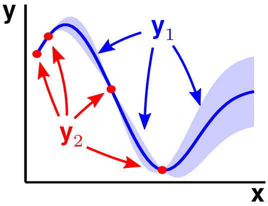
;

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} A & B \\ B^{\top} & C \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

;



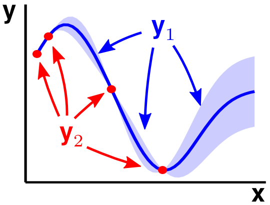
Mathematical Foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\mathbf{y}_1 ; \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\top)$$

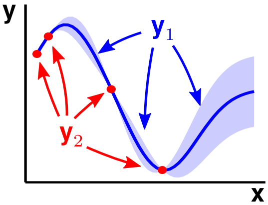


Mathematical Foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



$$\implies p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\ \mathbf{y}_1\ ;\ \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\top)$$

predictive mean

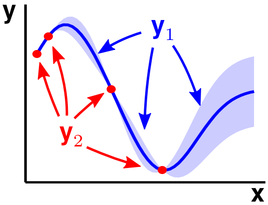
$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

Mathematical Foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\mathbf{y}_1 ; \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\top)$$

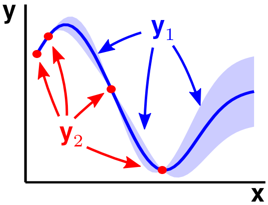
predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

$$= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2$$

Mathematical Foundations: Prediction

Q4. How do we make predictions?



$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\mathbf{y}_1 ; \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\top)$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

$$= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2$$

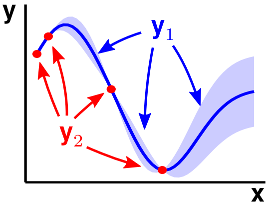
$$= \mathsf{W}\mathbf{y}_2$$

Mathematical Foundations: Prediction

Q4. How do we make predictions?

$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{B}^\top & \mathsf{C} \end{bmatrix}\right)$$

$$p(\mathbf{y}_1|\mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$



$$\Longrightarrow \quad p(\mathbf{y}_1|\mathbf{y}_2) = \mathcal{N}(\mathbf{y}_1; \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b}), \mathsf{A} - \mathsf{B}\mathsf{C}^{-1}\mathsf{B}^\top)$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathsf{B}\mathsf{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

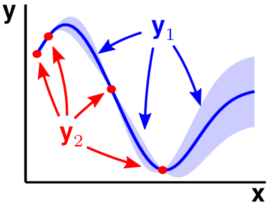
$$= \mathsf{B}\mathsf{C}^{-1}\mathbf{y}_2$$

$$= \mathsf{W}\mathbf{y}_2$$

linear in the data

Mathematical Foundations: Prediction

Q4. How do we make predictions?



$$p(\mathbf{y}_1, \mathbf{y}_2) = \mathcal{N}\left(\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} ; \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \begin{bmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{B}^\top & \mathrm{C} \end{bmatrix} \right)$$

$$p(\mathbf{y}_1 | \mathbf{y}_2) = \frac{p(\mathbf{y}_1, \mathbf{y}_2)}{p(\mathbf{y}_2)}$$

$$\Longrightarrow \quad p(\mathbf{y}_1 | \mathbf{y}_2) = \mathcal{N}(\ \mathbf{y}_1 ; \underline{\mathbf{a} + \mathrm{B}\mathrm{C}^{-1}(\mathbf{y}_2 - \mathbf{b})}, \underline{\mathrm{A} - \mathrm{B}\mathrm{C}^{-1}\mathrm{B}^\top})$$

predictive mean

$$\mu_{\mathbf{y}_1|\mathbf{y}_2} = \mathbf{a} + \mathrm{B}\mathrm{C}^{-1}(\mathbf{y}_2 - \mathbf{b})$$

$$= \mathrm{B}\mathrm{C}^{-1}\mathbf{y}_2$$

$$= \mathrm{W}\mathbf{y}_2$$

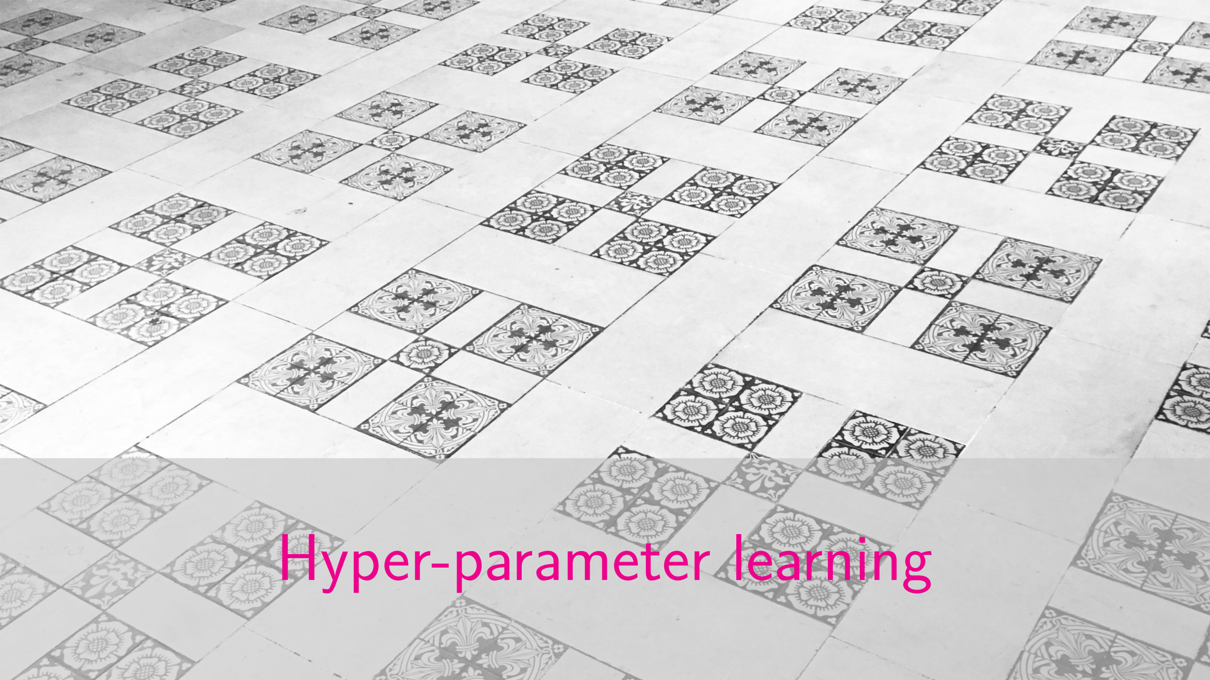
linear in the data

predictive covariance

$$\Sigma_{\mathbf{y}_1|\mathbf{y}_2} = \mathrm{A} - \mathrm{B}\mathrm{C}^{-1}\mathrm{B}^\top$$

predictive uncertainty = prior uncertainty − reduction in uncertainty

predictions more confident than prior



Hyper-parameter learning

What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

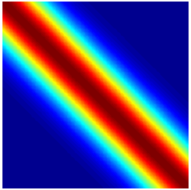
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

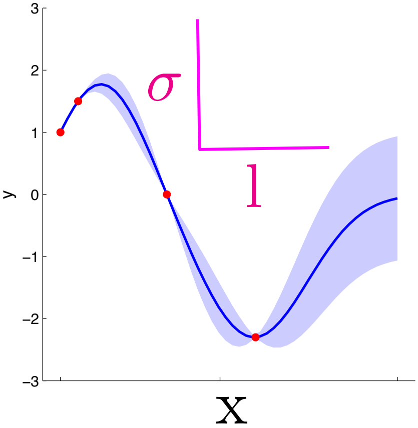
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

vertical-scale horizontal-scale

$\Sigma =$





Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

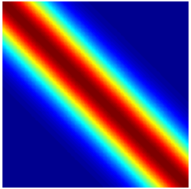
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

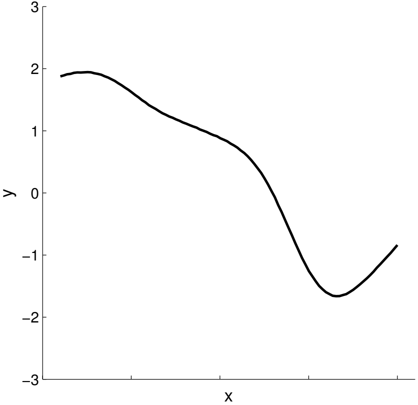
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

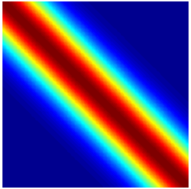
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

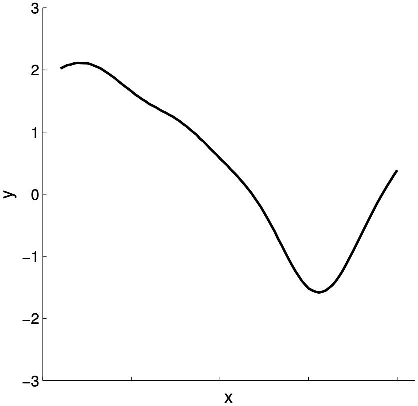
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

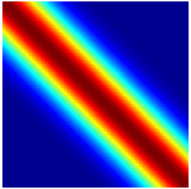
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

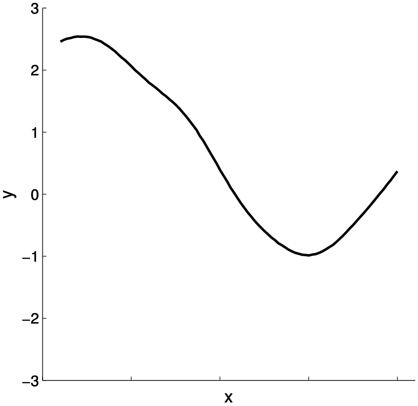
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

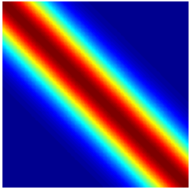
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

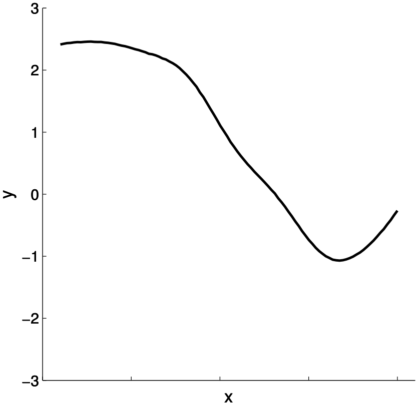
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

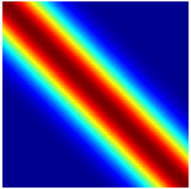
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

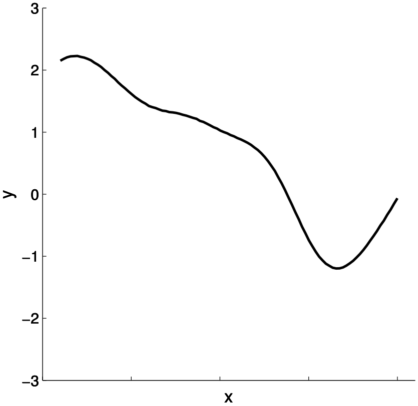
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

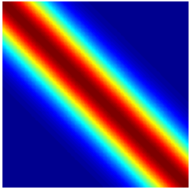
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

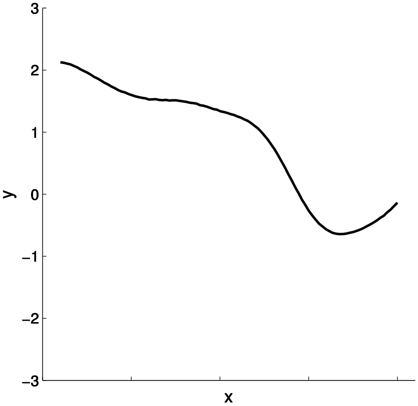
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

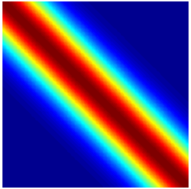
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

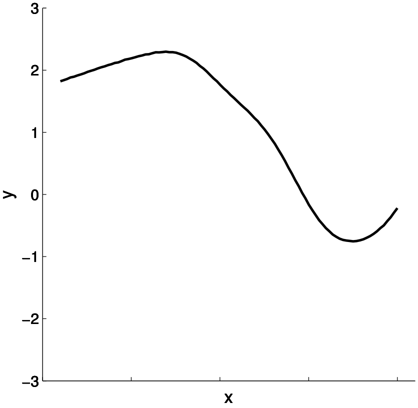
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

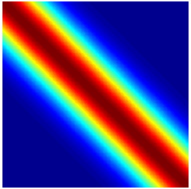
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

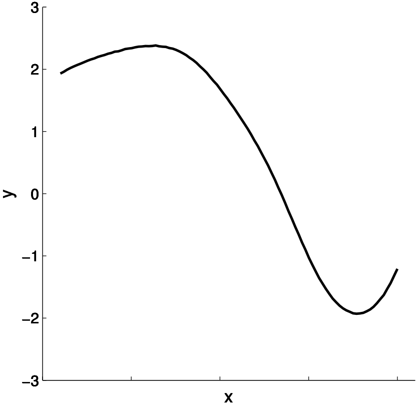
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

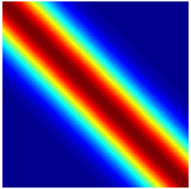
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

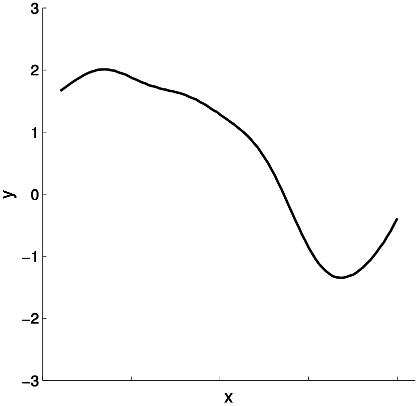
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

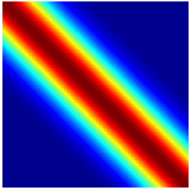
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

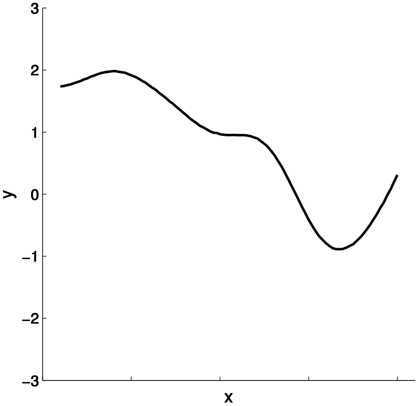
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

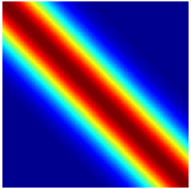
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

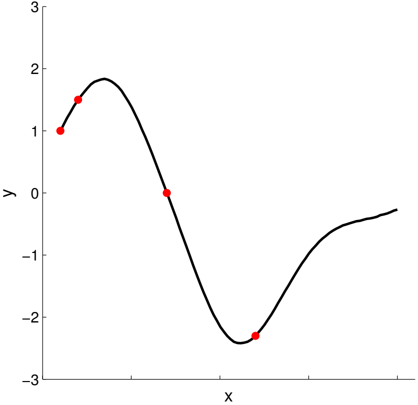
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

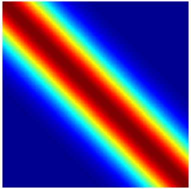
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

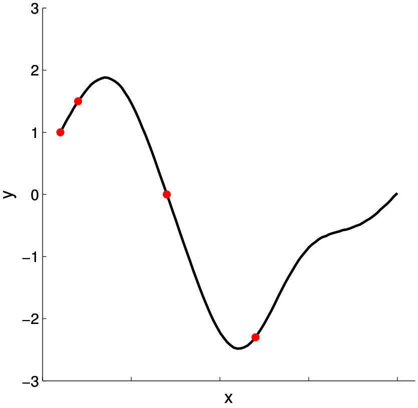
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

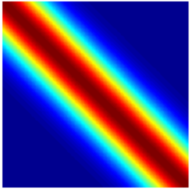
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

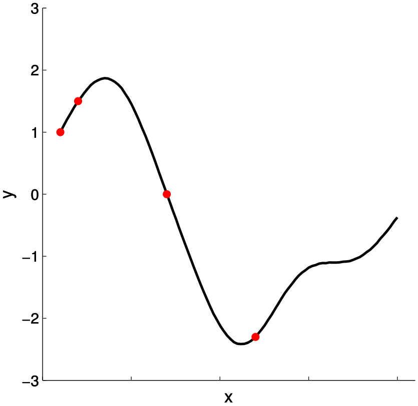
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

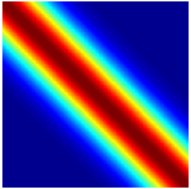
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

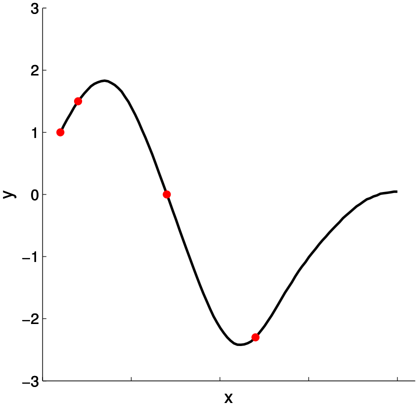
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

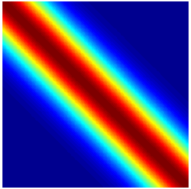
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

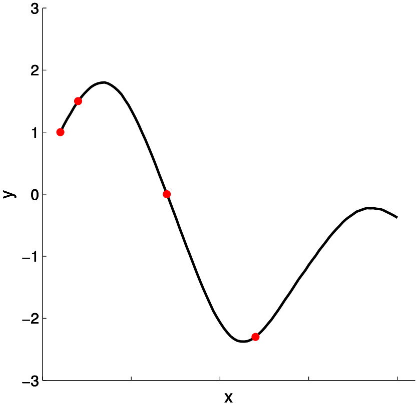
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

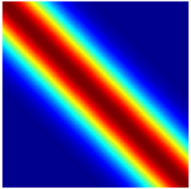
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

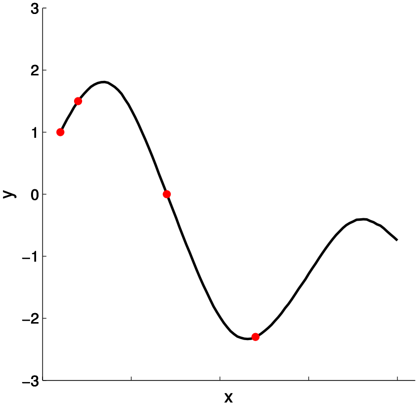
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2l^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

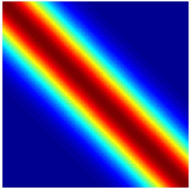
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

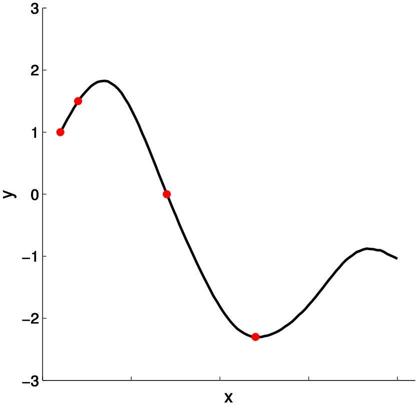
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

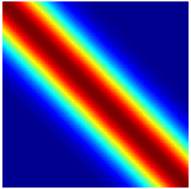
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

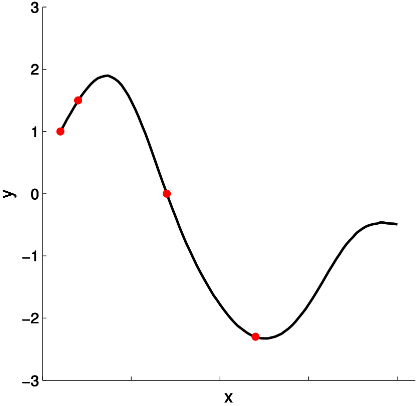
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

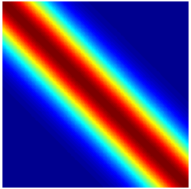
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

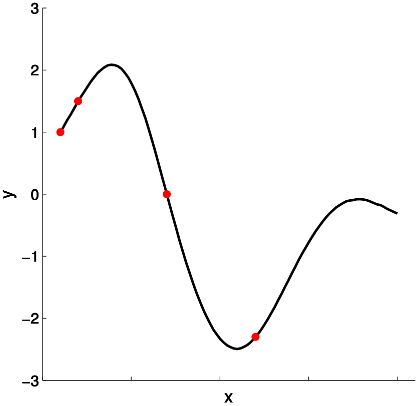
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

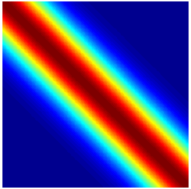
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

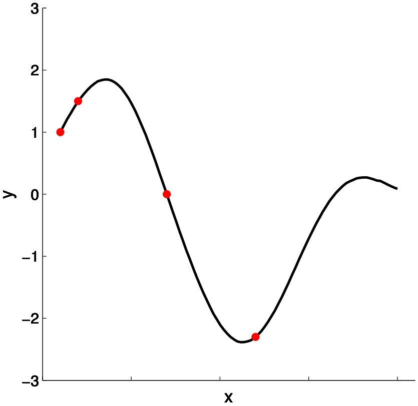
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

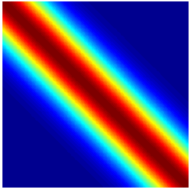
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

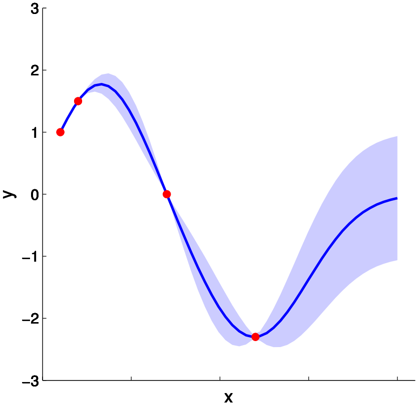
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

medium horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

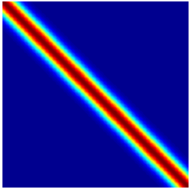
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

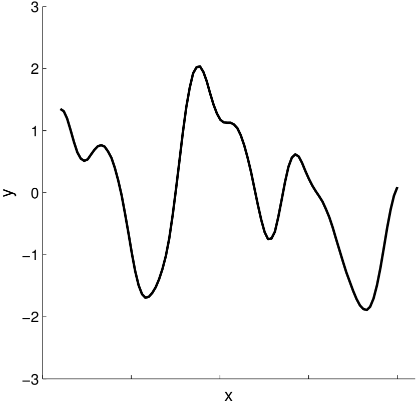
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

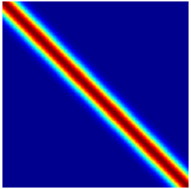
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

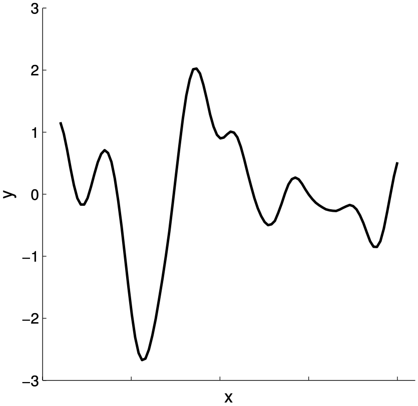
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

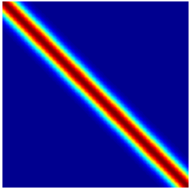
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

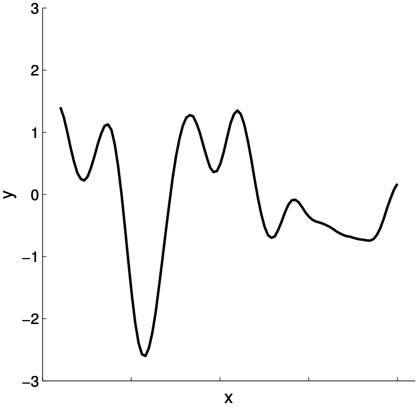
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

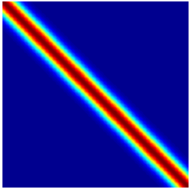
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

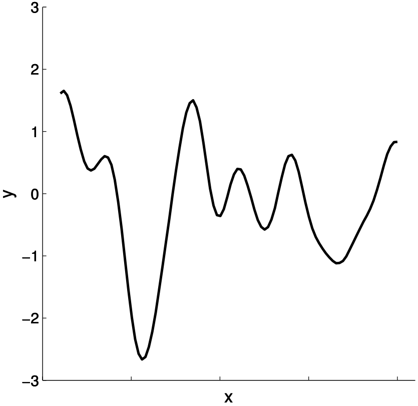
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

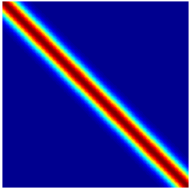
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

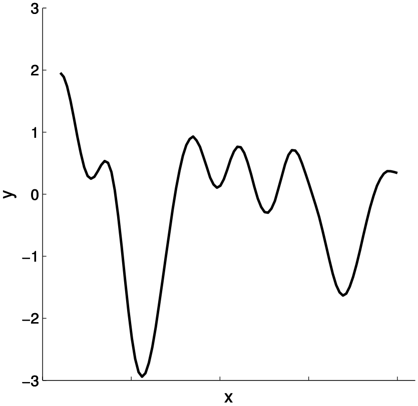
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

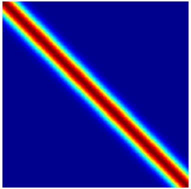
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

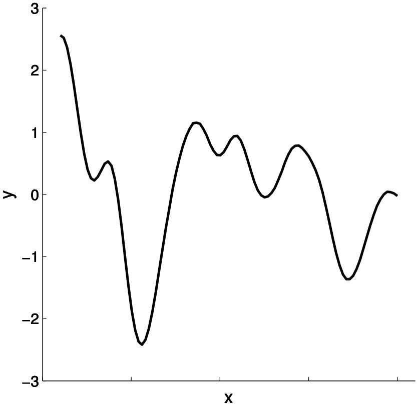
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

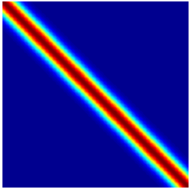
Non-parametric (∞-parametric)

$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

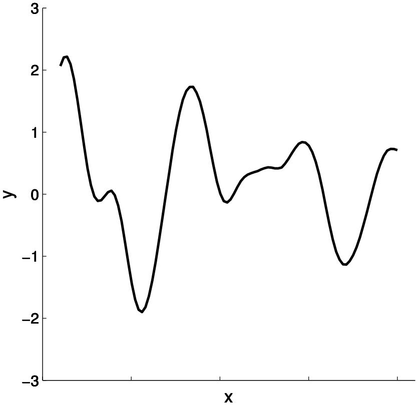
$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$$\Sigma =$$



short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

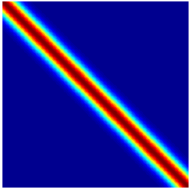
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

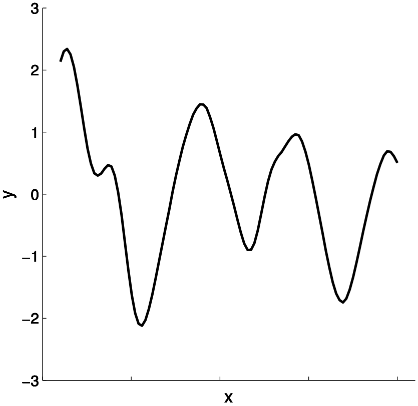
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

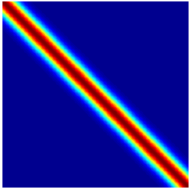
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

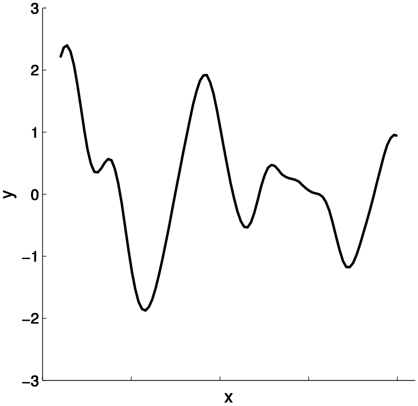
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

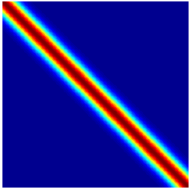
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

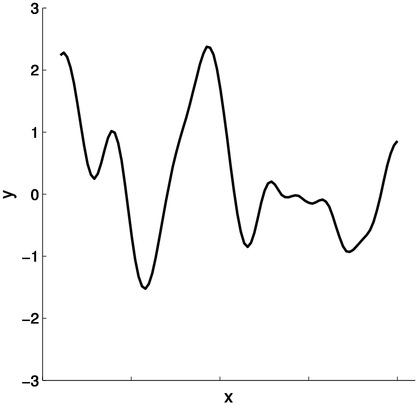
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

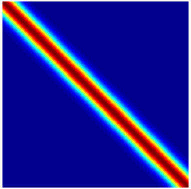
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

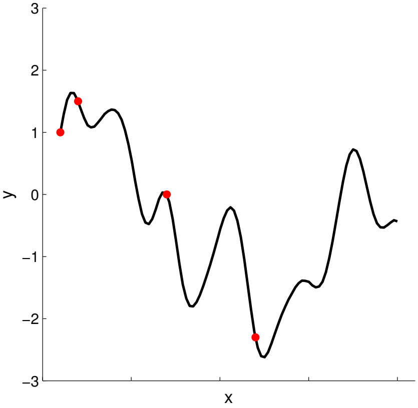
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

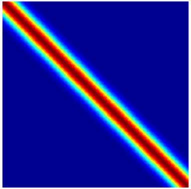
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

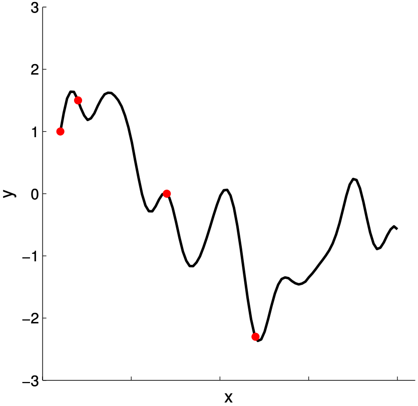
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

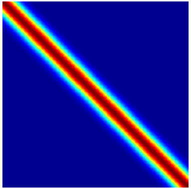
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

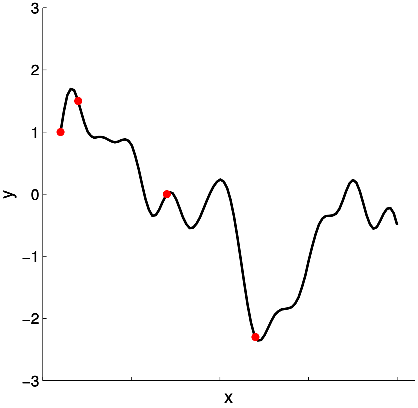
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

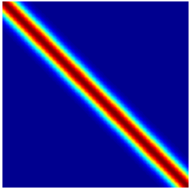
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

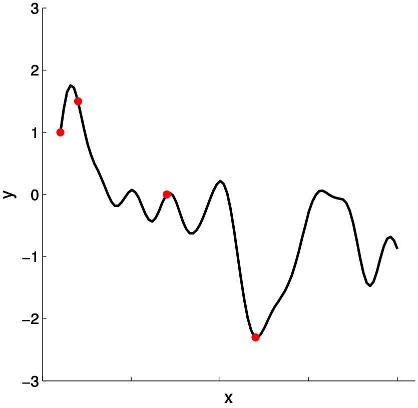
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

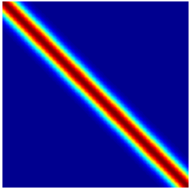
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

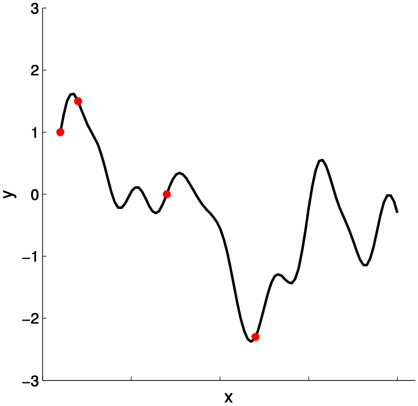
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

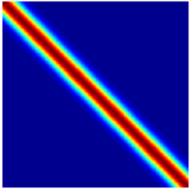
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

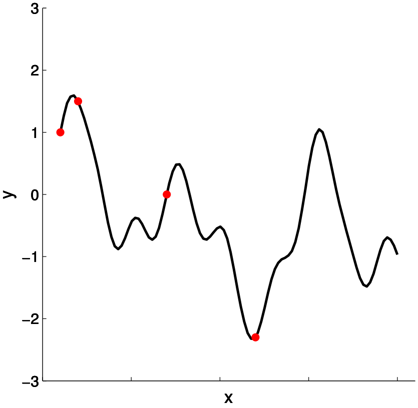
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

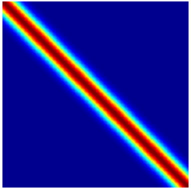
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

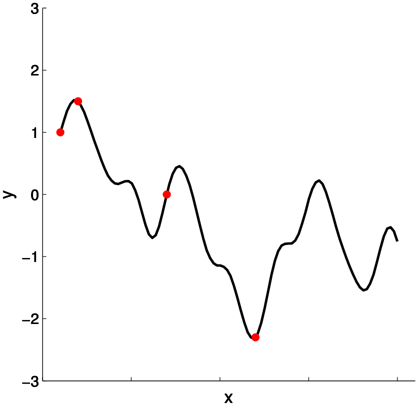
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

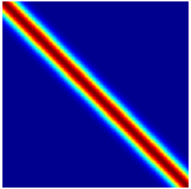
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

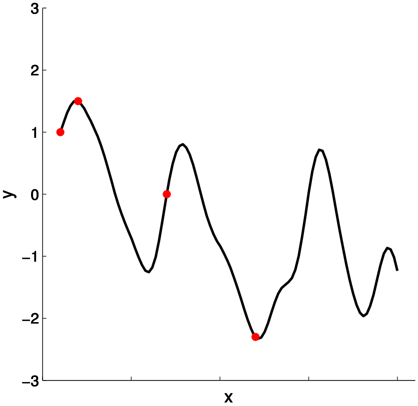
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

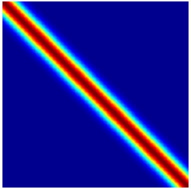
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

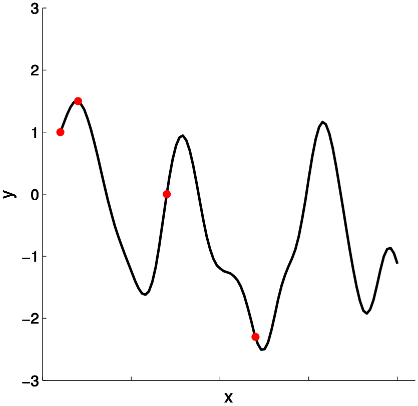
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

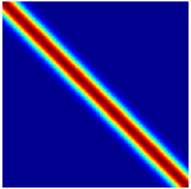
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

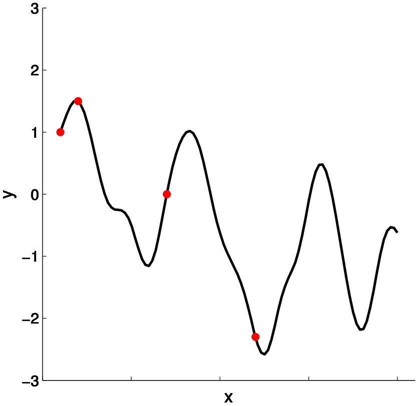
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

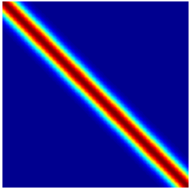
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

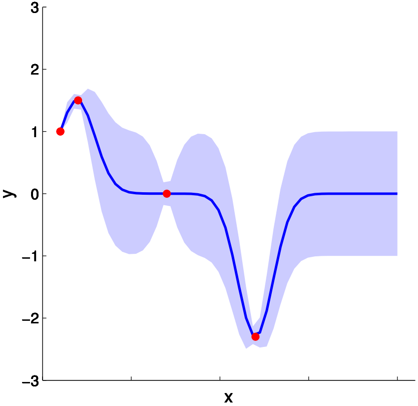
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

short horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

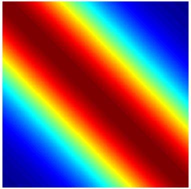
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

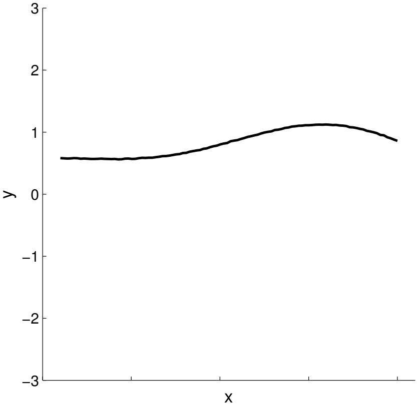
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

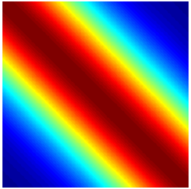
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

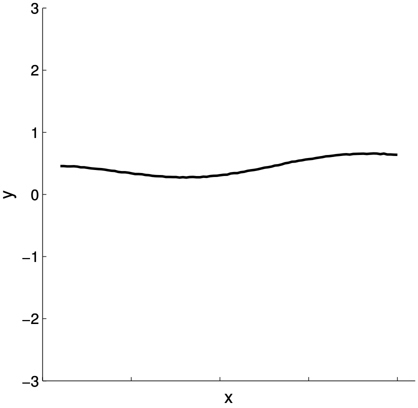
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

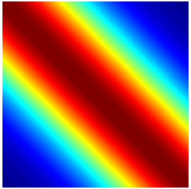
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

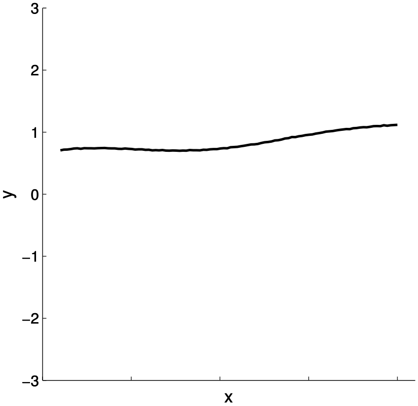
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

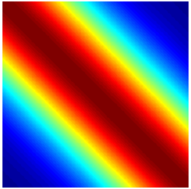
Non-parametric (∞-parametric)

$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

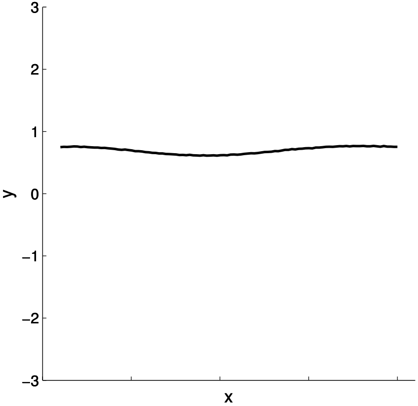
$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$



long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

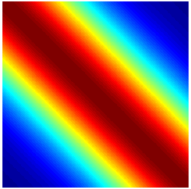
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

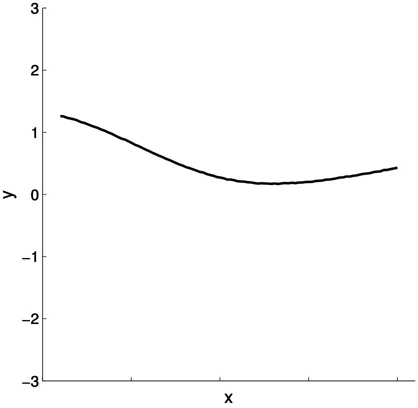
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

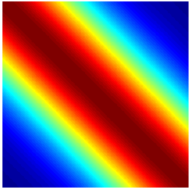
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

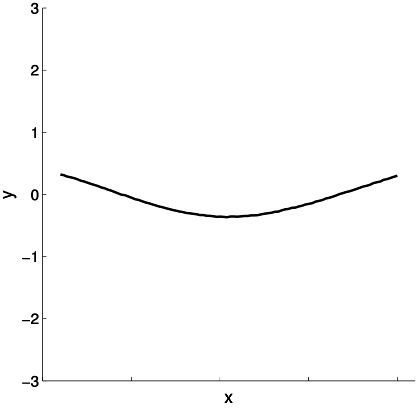
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

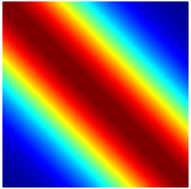
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

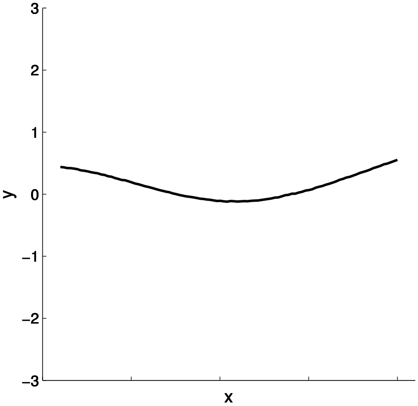
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

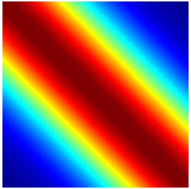
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

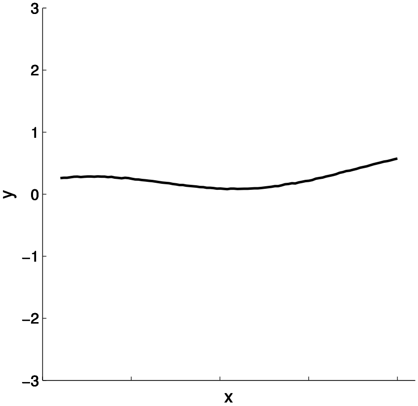
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

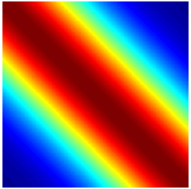
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

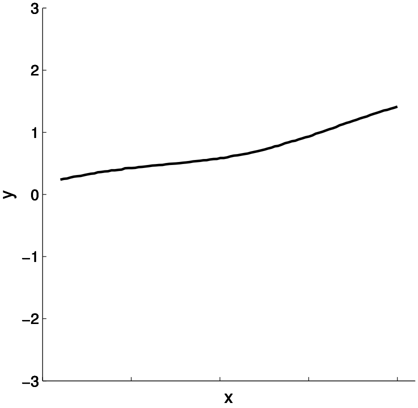
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_\mathsf{y}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_\mathsf{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

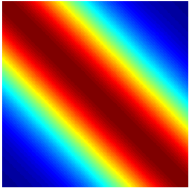
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

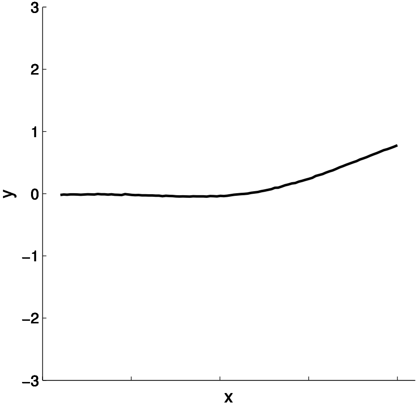
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

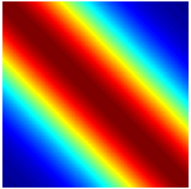
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

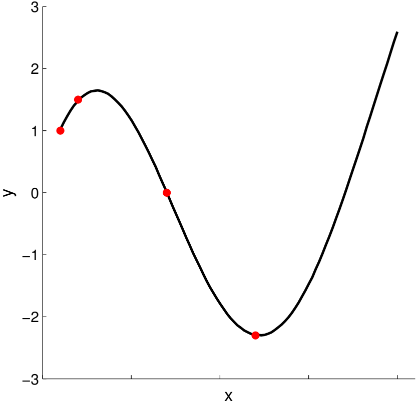
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

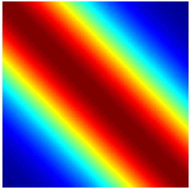
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

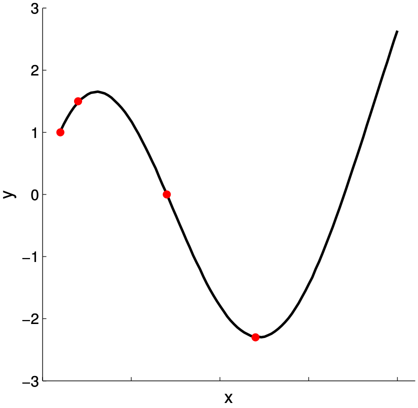
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

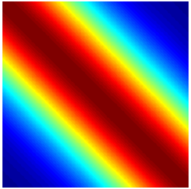
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

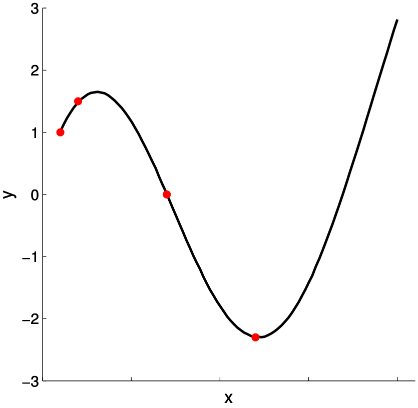
$$p(\mathsf{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathsf{x}_1, \mathsf{x}_2) = \mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) + \mathsf{I}\sigma_{\mathsf{y}}^2$$

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 e^{-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathsf{y}(\mathsf{x}) = f(\mathsf{x}; \theta) + \sigma_{\mathsf{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

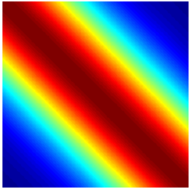
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

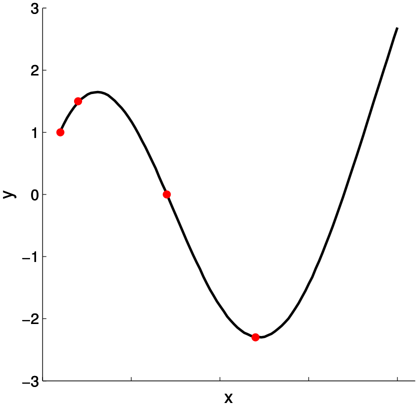
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

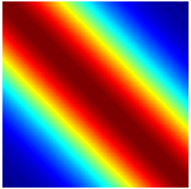
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

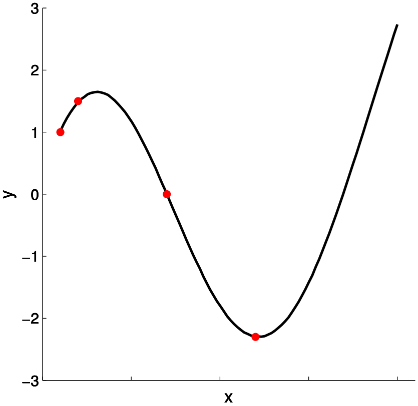
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 \mathrm{e}^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

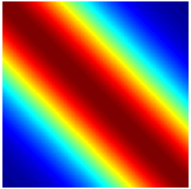
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

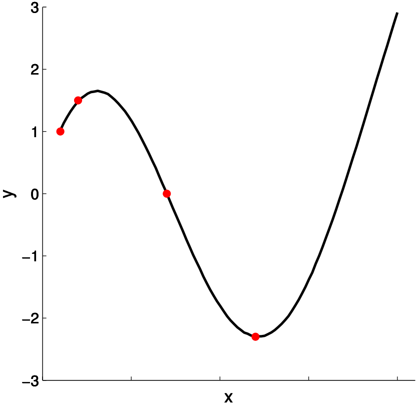
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

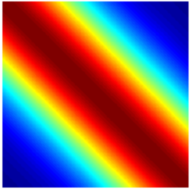
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

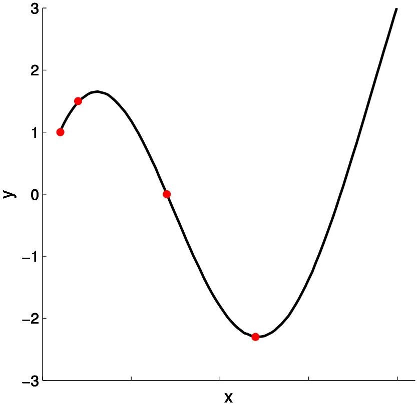
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

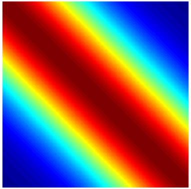
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

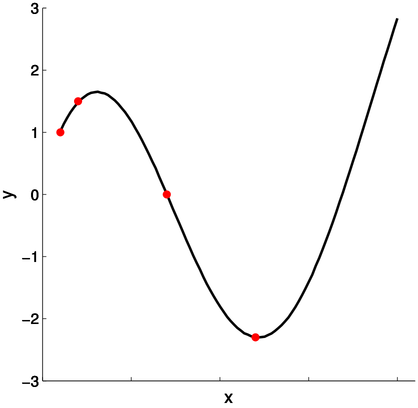
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

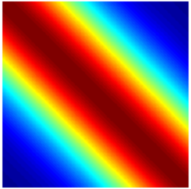
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

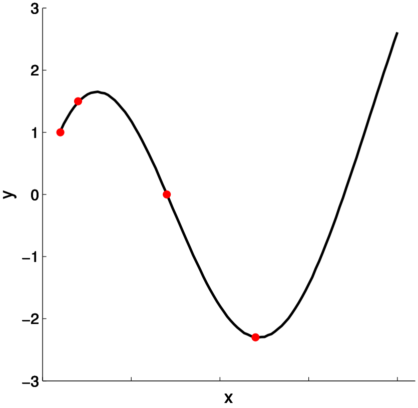
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_\mathrm{y}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_\mathrm{y}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

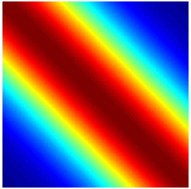
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

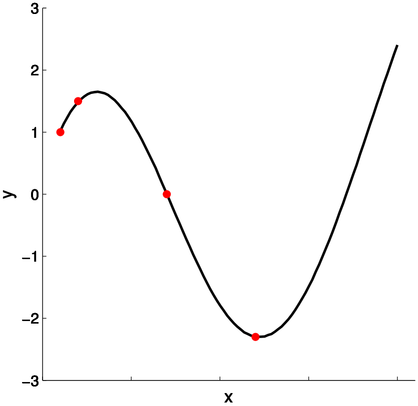
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

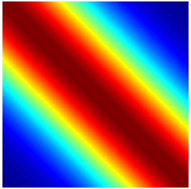
What effect do the hyper-parameters have?

Non-parametric (∞-parametric)

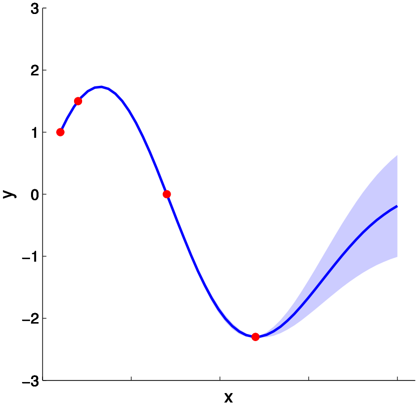
$$p(\mathrm{y}|\theta) = \mathcal{N}(y; 0, \Sigma)$$

$$\Sigma(\mathrm{x}_1, \mathrm{x}_2) = \mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{I}\sigma_{\mathrm{y}}^2$$

$$\mathrm{K}(\mathrm{x}_1, \mathrm{x}_2) = \sigma^2 e^{-\frac{1}{2\mathrm{l}^2}(\mathrm{x}_1 - \mathrm{x}_2)^2}$$

$\Sigma =$ 

long horizontal length-scale



Parametric model

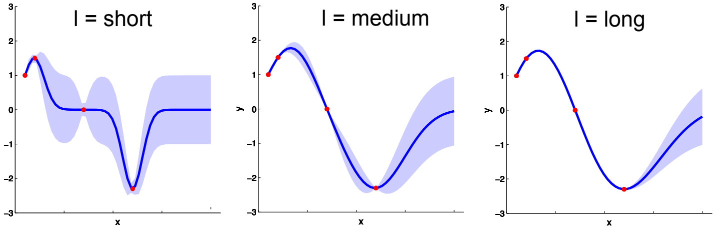
$$\mathrm{y}(\mathrm{x}) = f(\mathrm{x}; \theta) + \sigma_{\mathrm{y}}\epsilon$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

What effect do the hyper-parameters have?

$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

- Hyper-parameters have a strong effect
  - l controls the horizontal length-scale
  - σ^2 controls the vertical scale of the data
- $\Longrightarrow$ need automatic learning of hyper-parameters from data



How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

what we know after seeing the data $\propto$ what the data tell us (likelihood) $\times$ what we knew before seeing the data (prior)

$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

How do we choose the hyper-parameters?

idea: use probability distributions to represent plausibility of hyper-parameters (uncertainty) given the data

what we know after seeing the data $\propto$ what the data tell us (likelihood) $\times$ what we knew before seeing the data (prior)

$$p(\theta|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|\theta)p(\theta)}{p(\mathbf{y}_{1:N})}$$ (Bayes' Rule)

$p(\mathbf{y}_{1:N}|\theta)$ = likelihood of the parameters

= how well did θ predict the data we observed

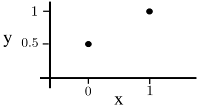
$$p(\mathbf{y}_{1:N}|\theta) = \frac{1}{\det(2\pi\Sigma(\theta))^{-1/2}} \exp\left(-\frac{1}{2}\mathbf{y}_{1:N}^{\top}\Sigma^{-1}(\theta)\mathbf{y}_{1:N}\right)$$

How do we choose the hyper-parameters?

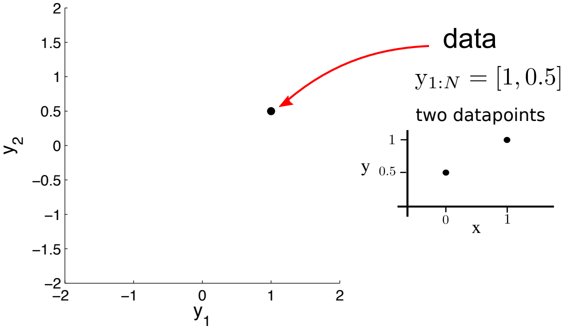
data

$$\mathrm{y}_{1:N} = [1, 0.5]$$

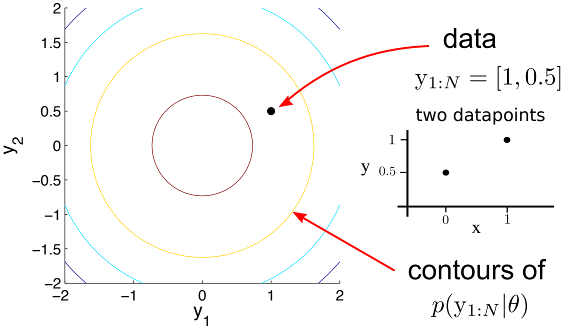
two datapoints



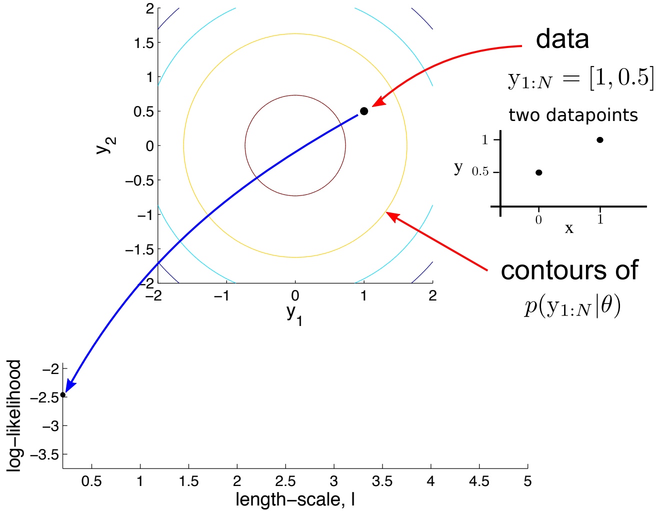
How do we choose the hyper-parameters?



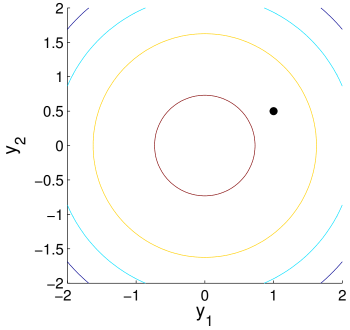
How do we choose the hyper-parameters?

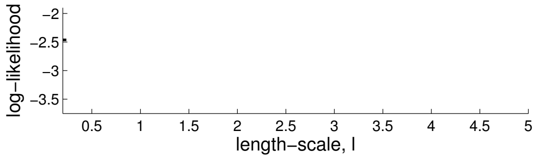


How do we choose the hyper-parameters?

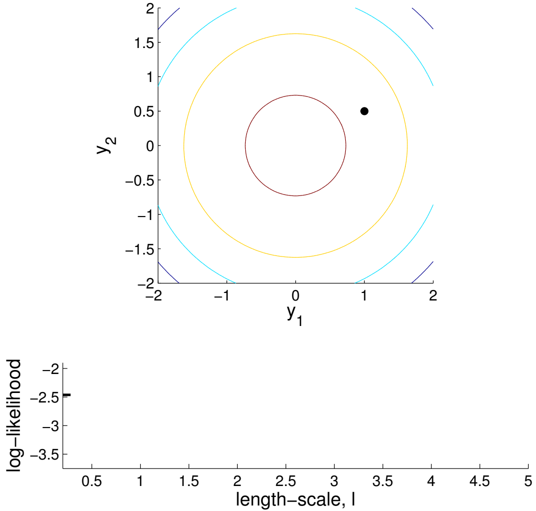


How do we choose the hyper-parameters?

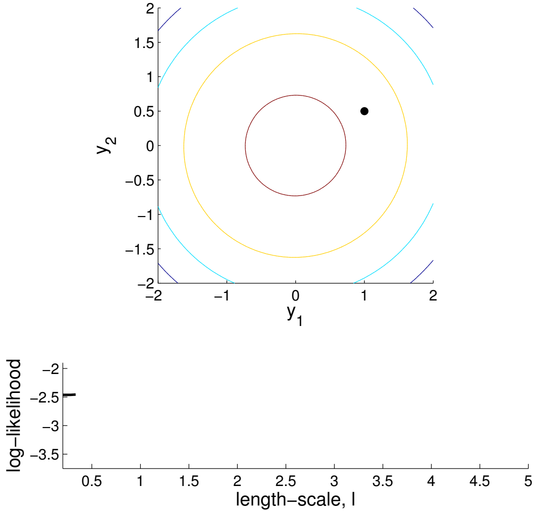




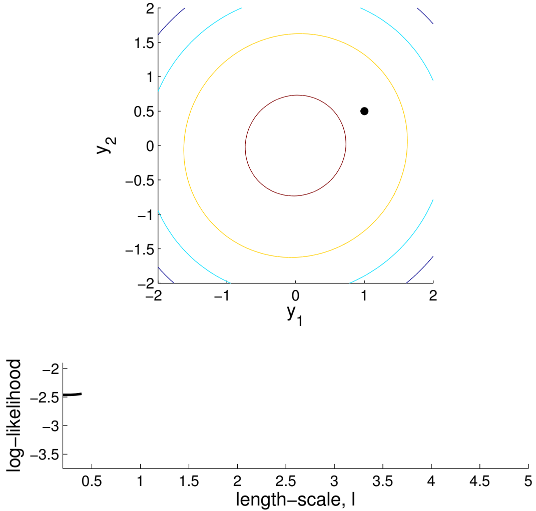
How do we choose the hyper-parameters?



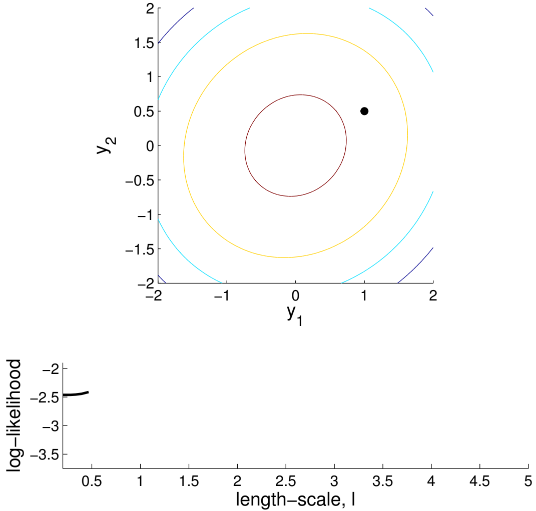
How do we choose the hyper-parameters?



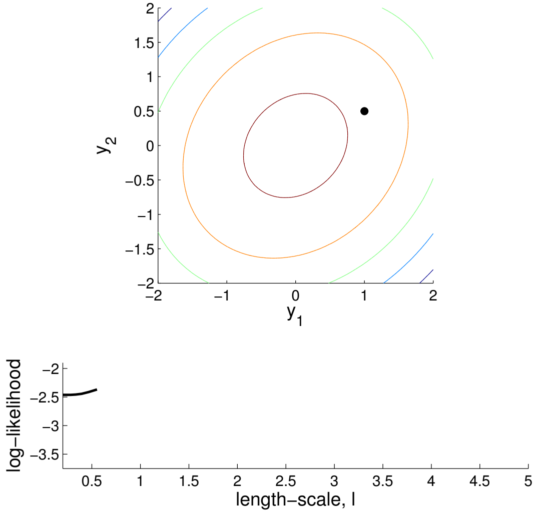
How do we choose the hyper-parameters?



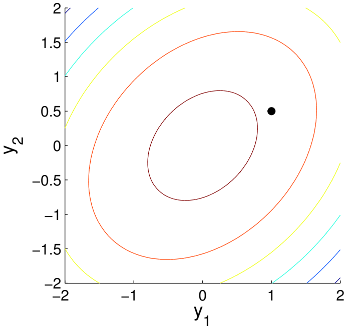
How do we choose the hyper-parameters?

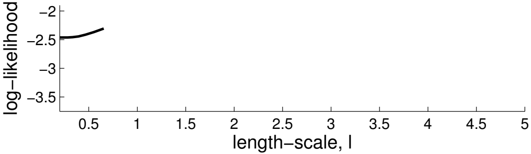


How do we choose the hyper-parameters?

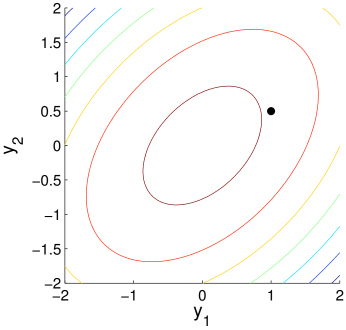


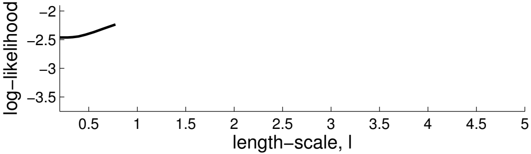
How do we choose the hyper-parameters?



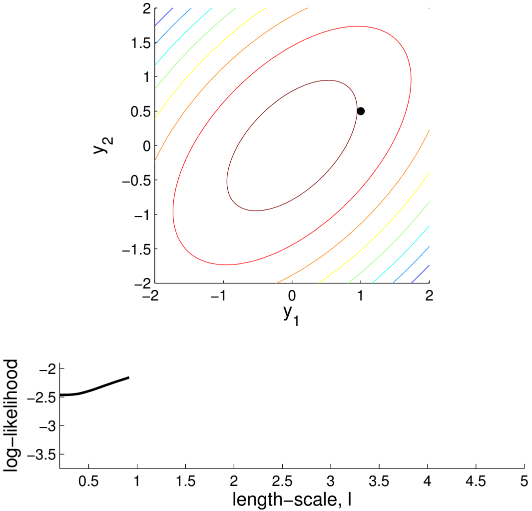


How do we choose the hyper-parameters?

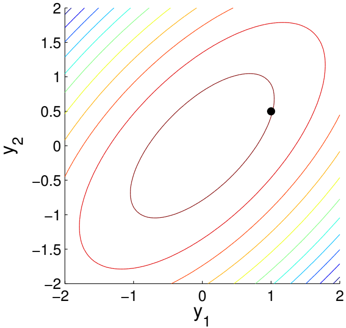


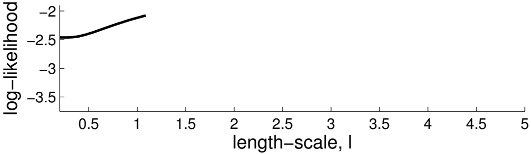


How do we choose the hyper-parameters?

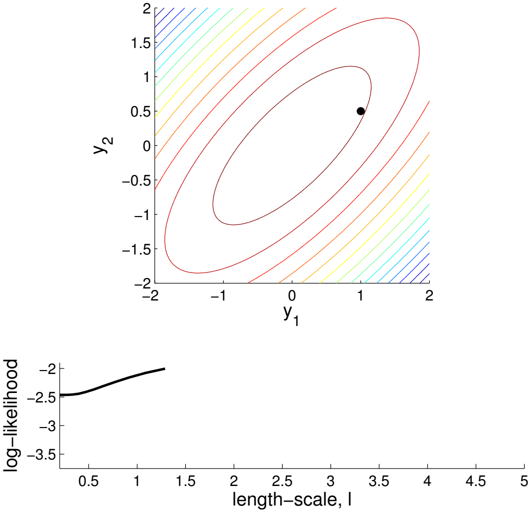


How do we choose the hyper-parameters?

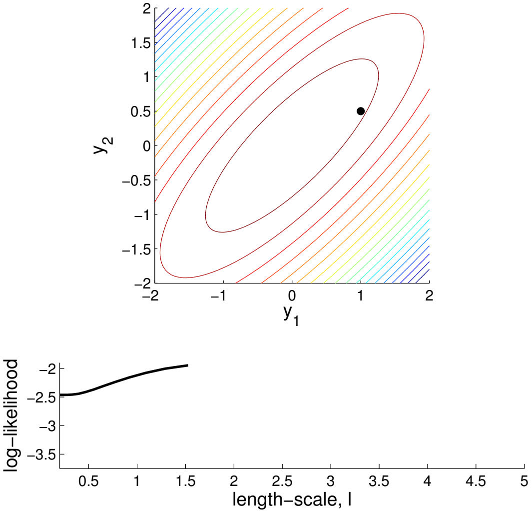




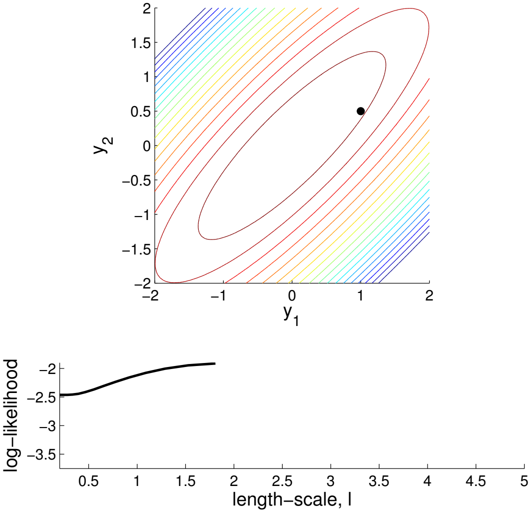
How do we choose the hyper-parameters?



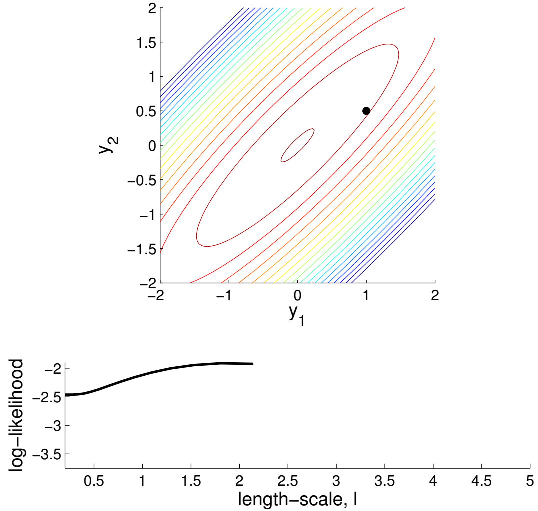
How do we choose the hyper-parameters?



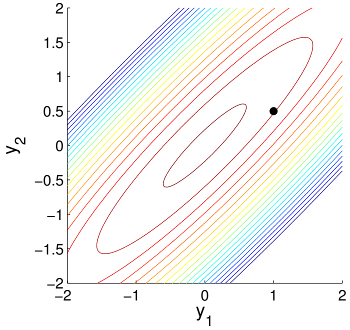
How do we choose the hyper-parameters?

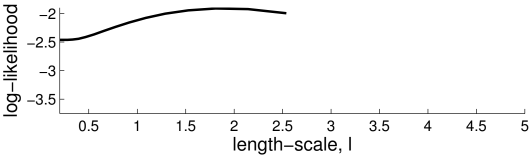


How do we choose the hyper-parameters?

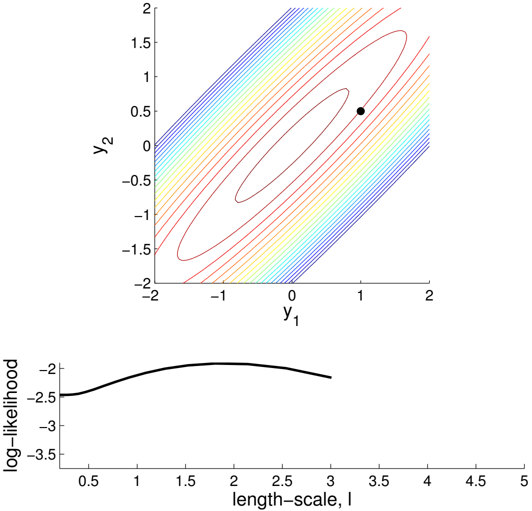


How do we choose the hyper-parameters?

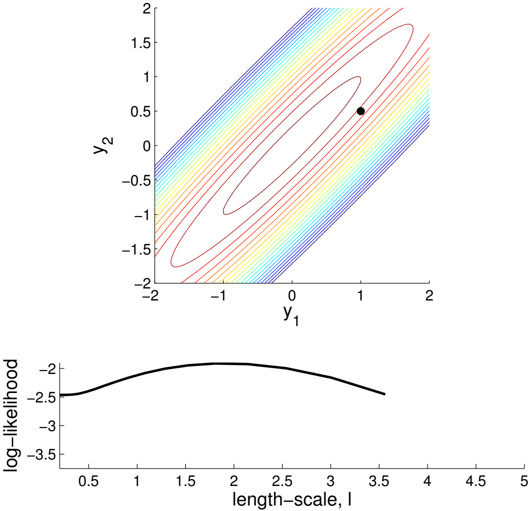




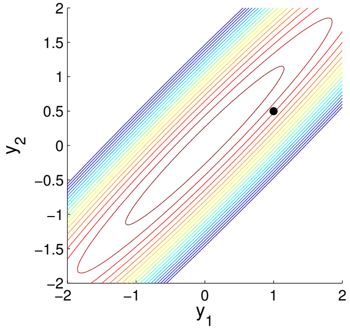
How do we choose the hyper-parameters?

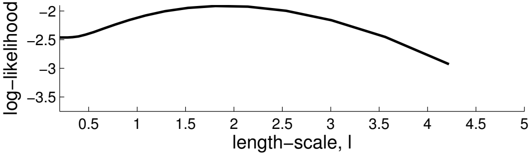


How do we choose the hyper-parameters?

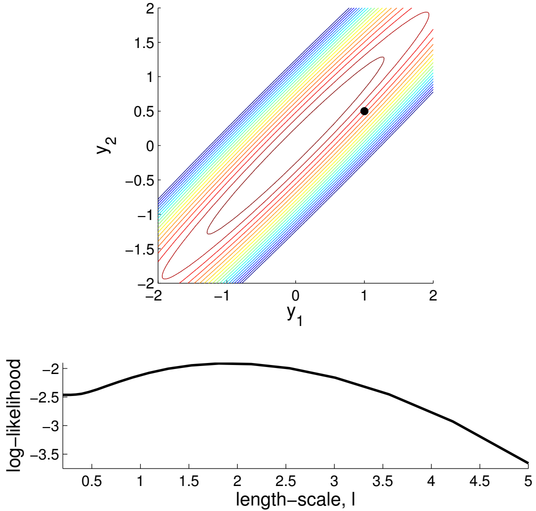


How do we choose the hyper-parameters?

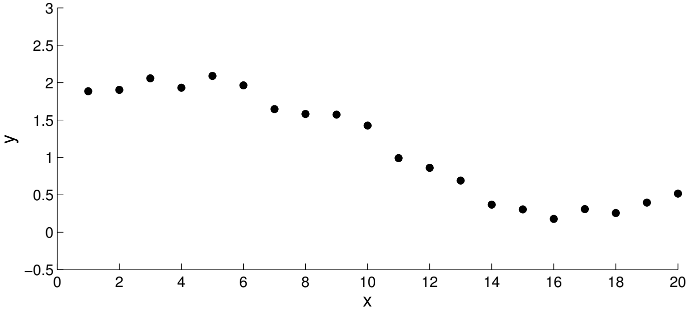




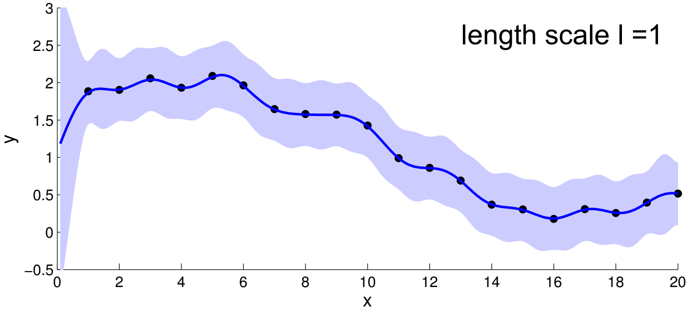
How do we choose the hyper-parameters?



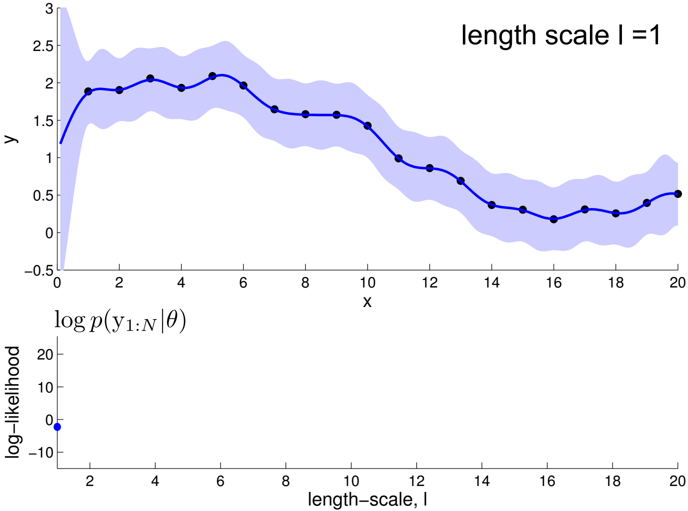
How do we choose the hyper-parameters?



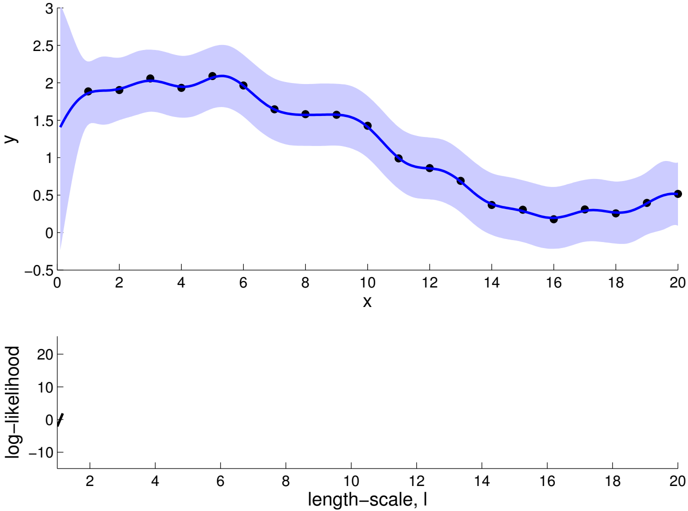
How do we choose the hyper-parameters?



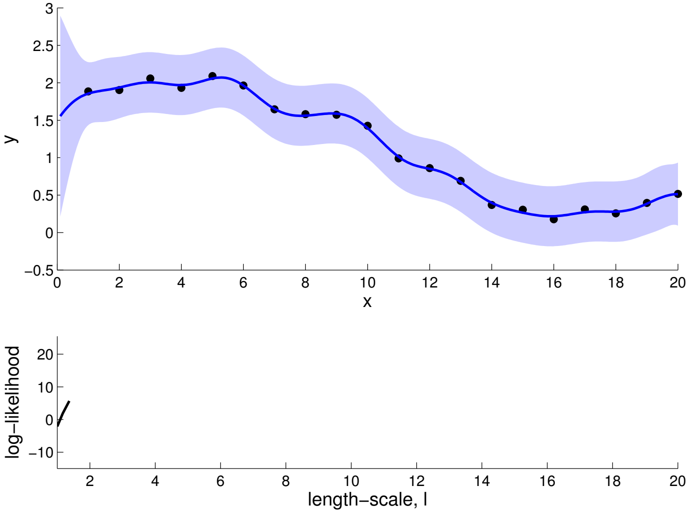
How do we choose the hyper-parameters?



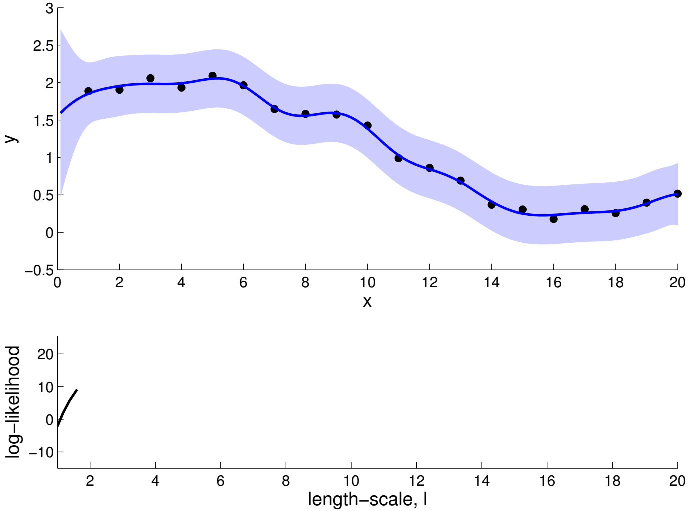
How do we choose the hyper-parameters?



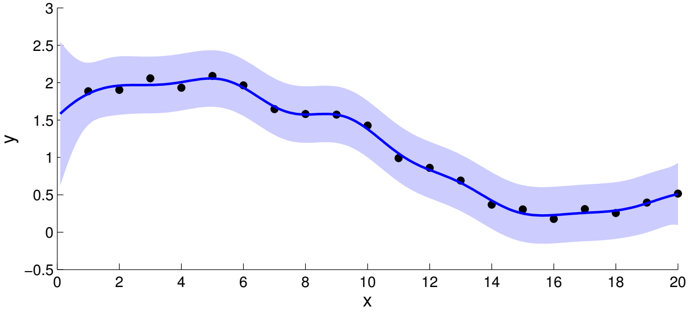
How do we choose the hyper-parameters?

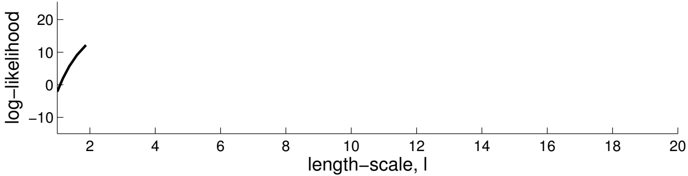


How do we choose the hyper-parameters?

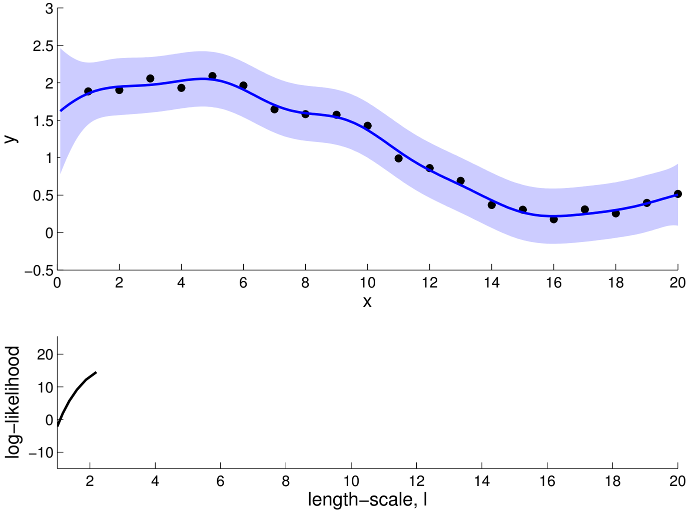


How do we choose the hyper-parameters?

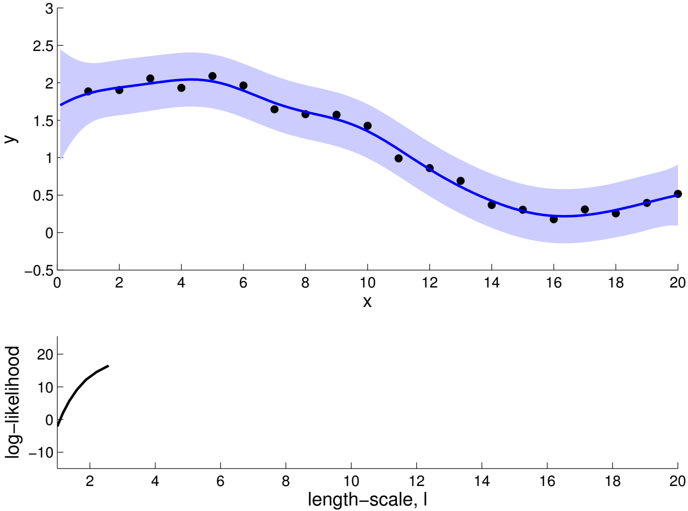




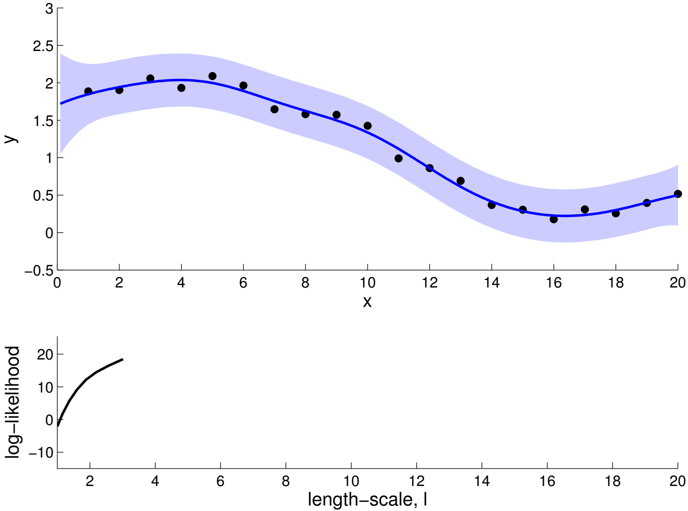
How do we choose the hyper-parameters?



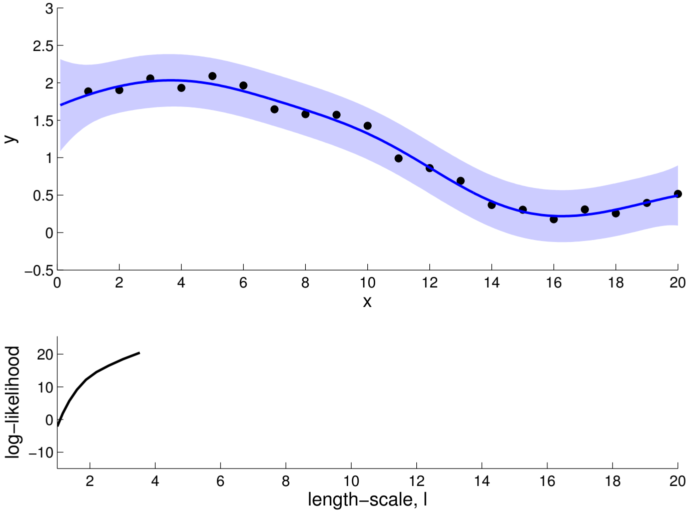
How do we choose the hyper-parameters?



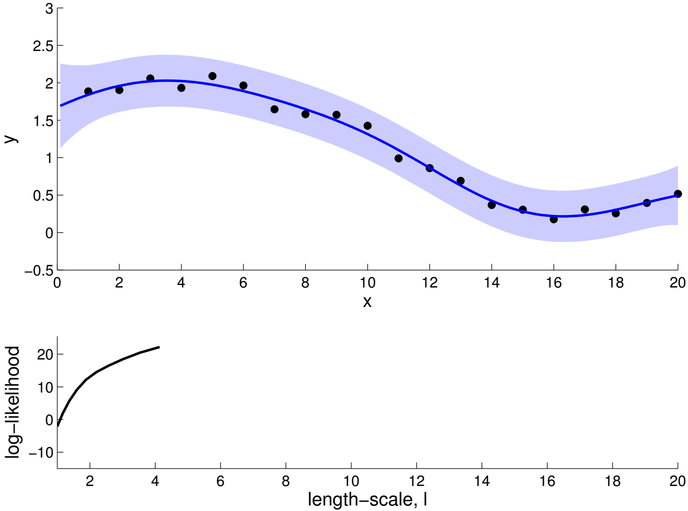
How do we choose the hyper-parameters?



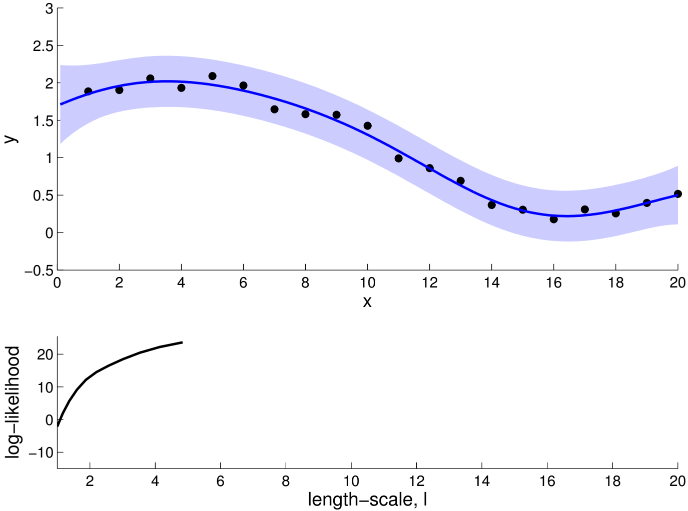
How do we choose the hyper-parameters?



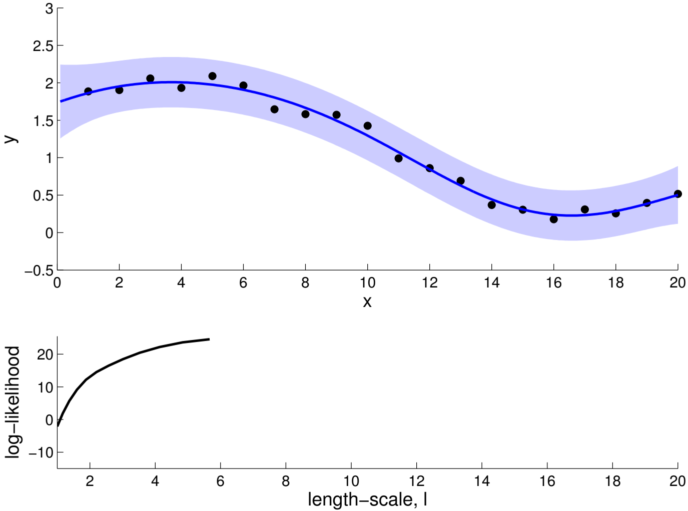
How do we choose the hyper-parameters?



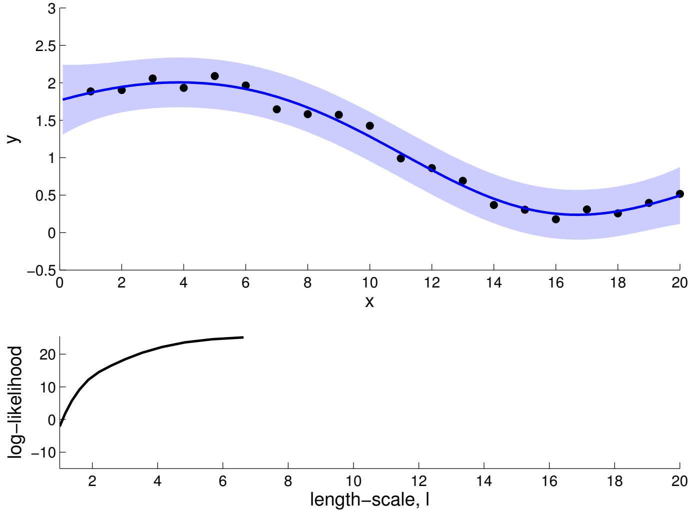
How do we choose the hyper-parameters?



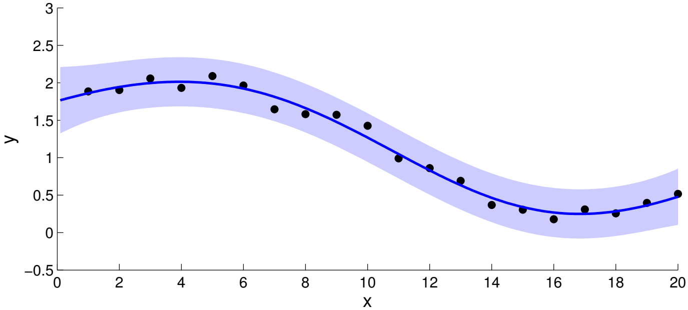
How do we choose the hyper-parameters?

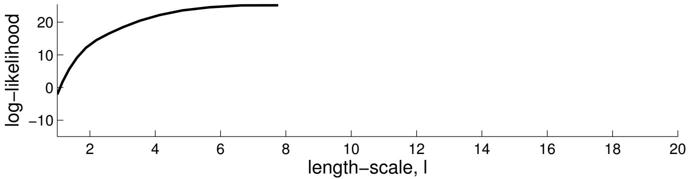


How do we choose the hyper-parameters?

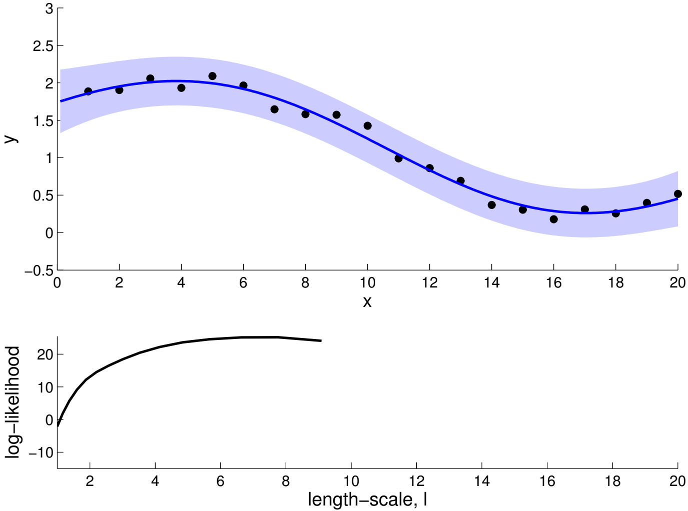


How do we choose the hyper-parameters?

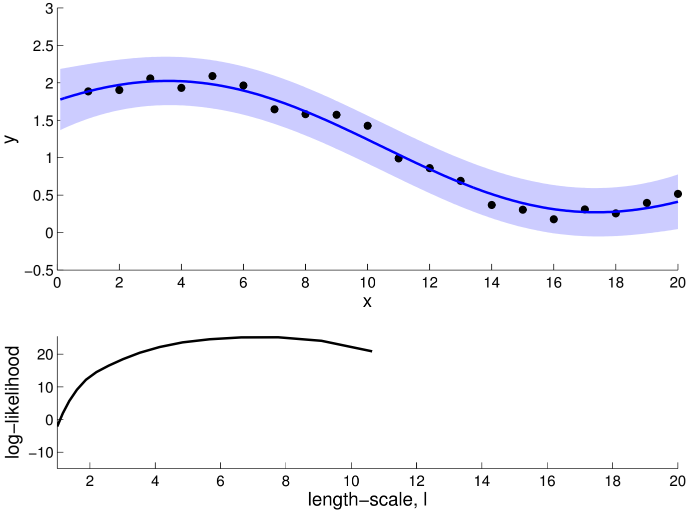




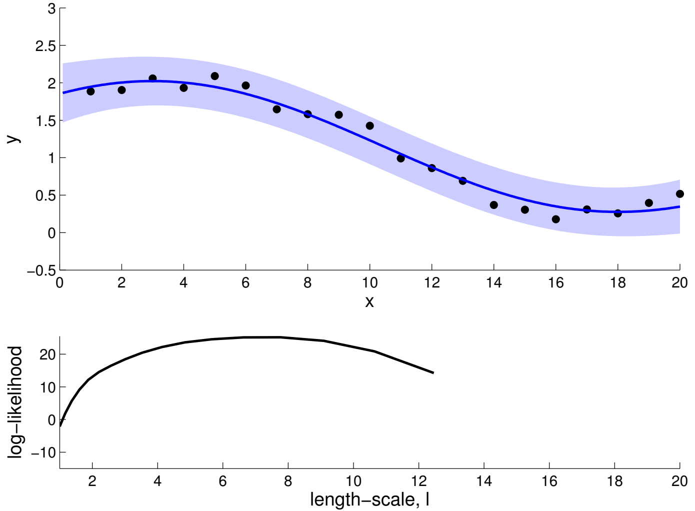
How do we choose the hyper-parameters?



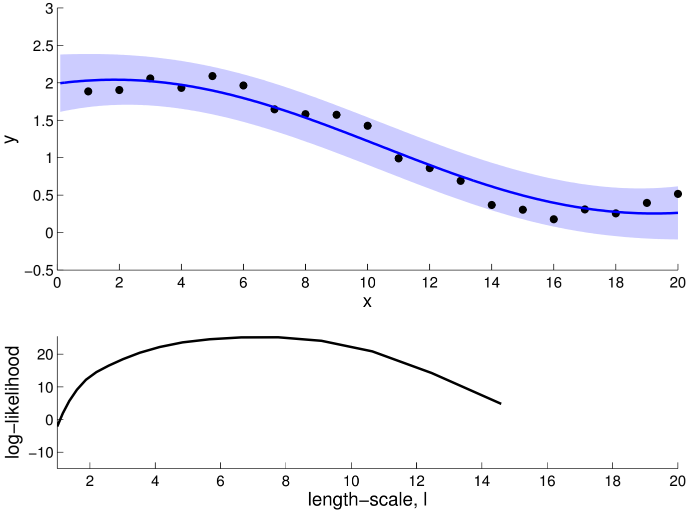
How do we choose the hyper-parameters?



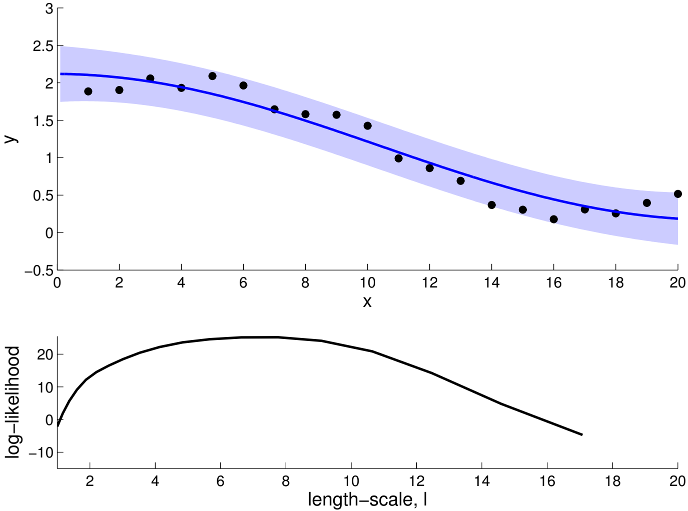
How do we choose the hyper-parameters?



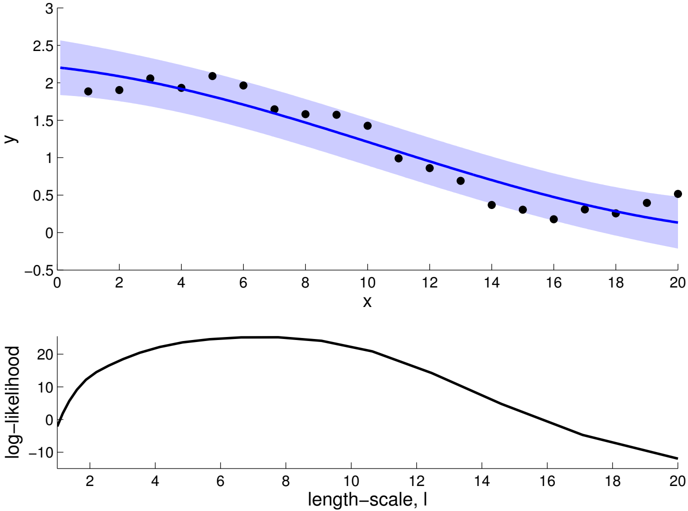
How do we choose the hyper-parameters?



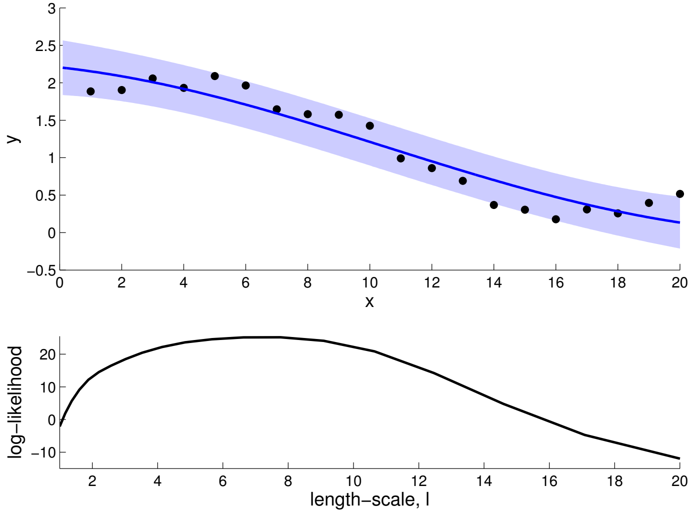
How do we choose the hyper-parameters?



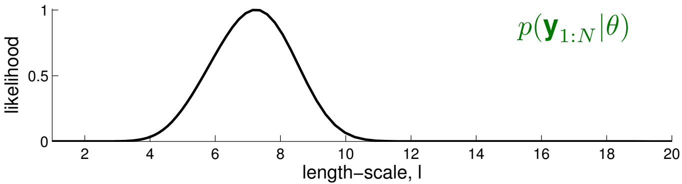
How do we choose the hyper-parameters?



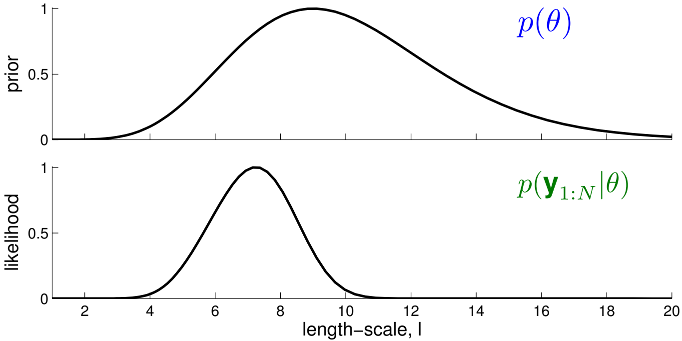
How do we choose the hyper-parameters?



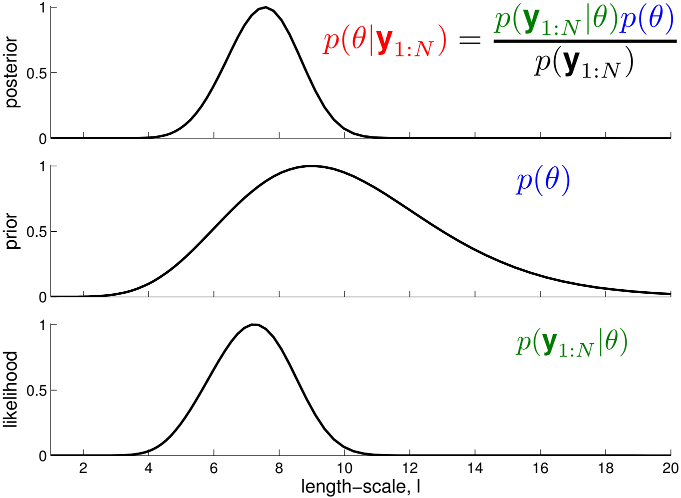
How do we choose the hyper-parameters?



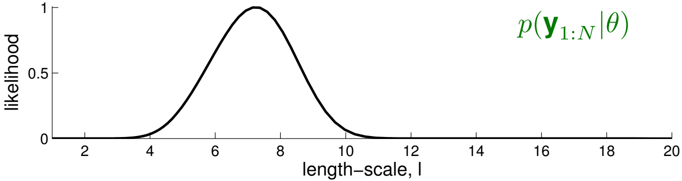
How do we choose the hyper-parameters?



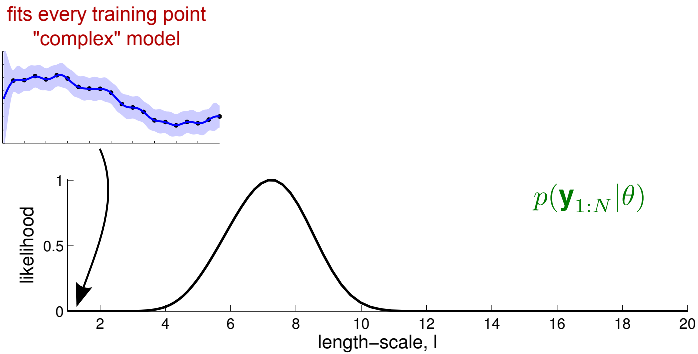
How do we choose the hyper-parameters?



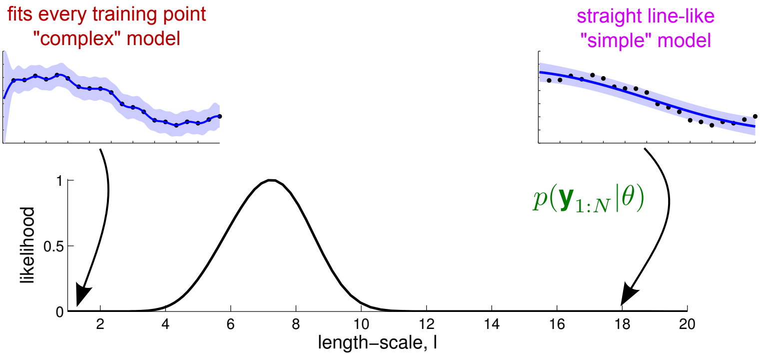
Why does Bayesian inference work?



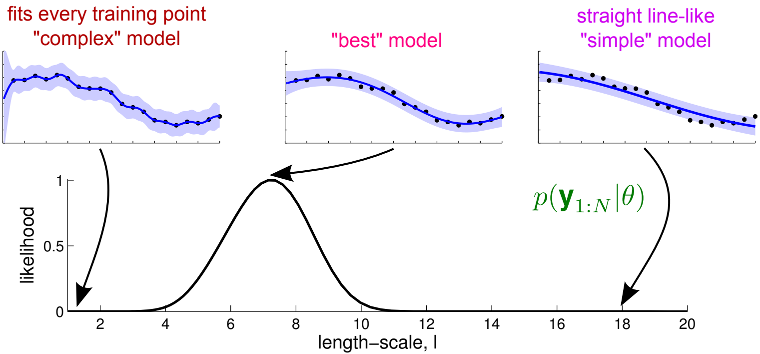
Why does Bayesian inference work?



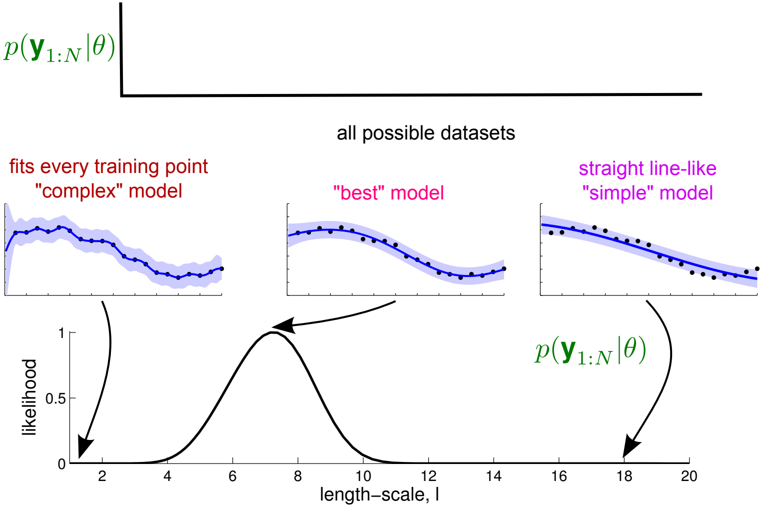
Why does Bayesian inference work?



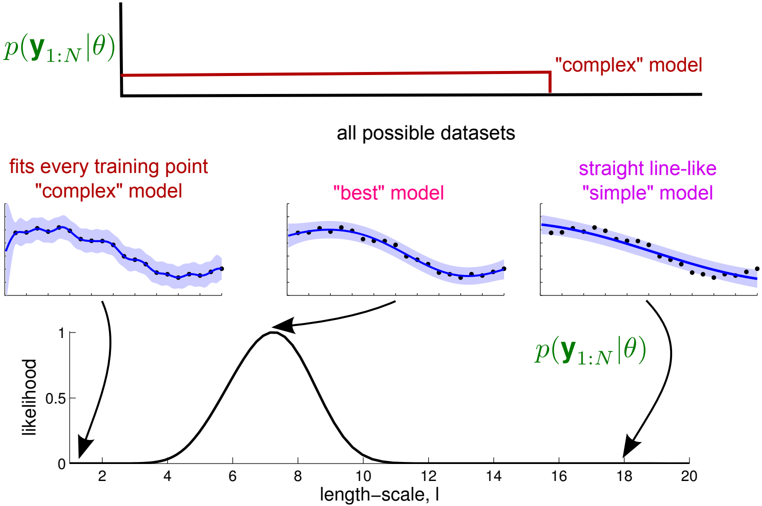
Why does Bayesian inference work?



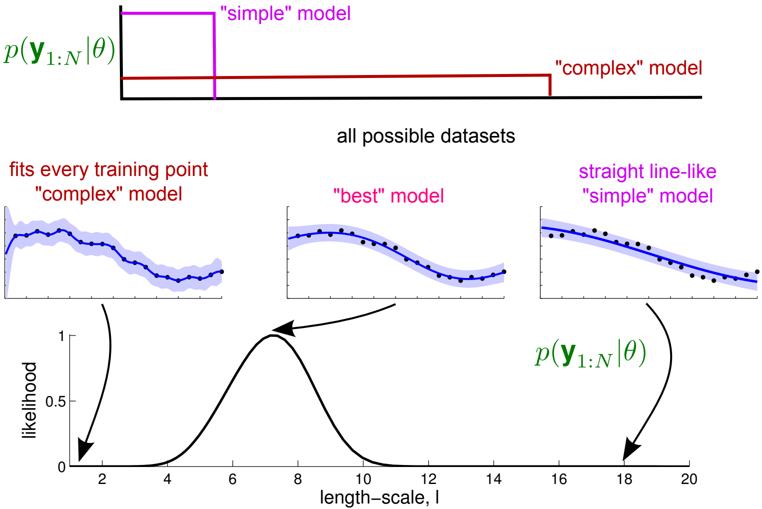
Why does Bayesian inference work?



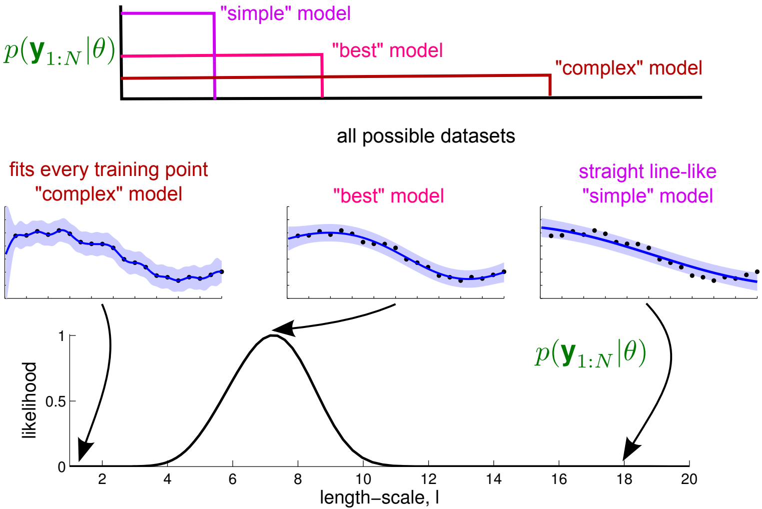
Why does Bayesian inference work?



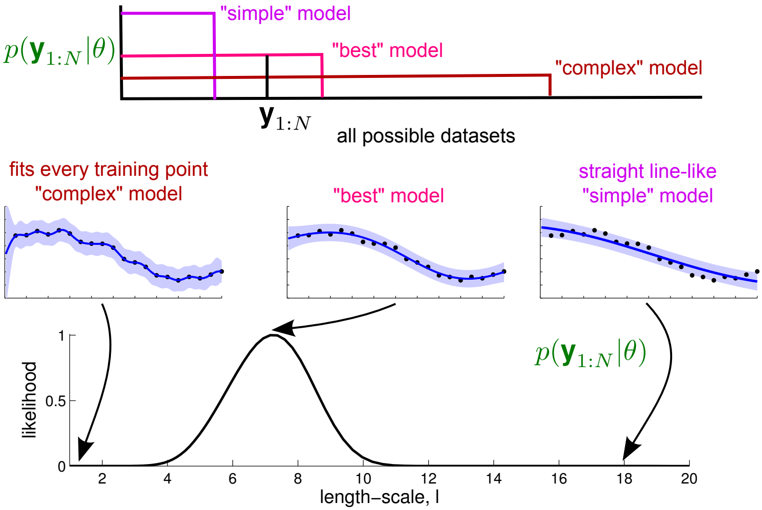
Why does Bayesian inference work?

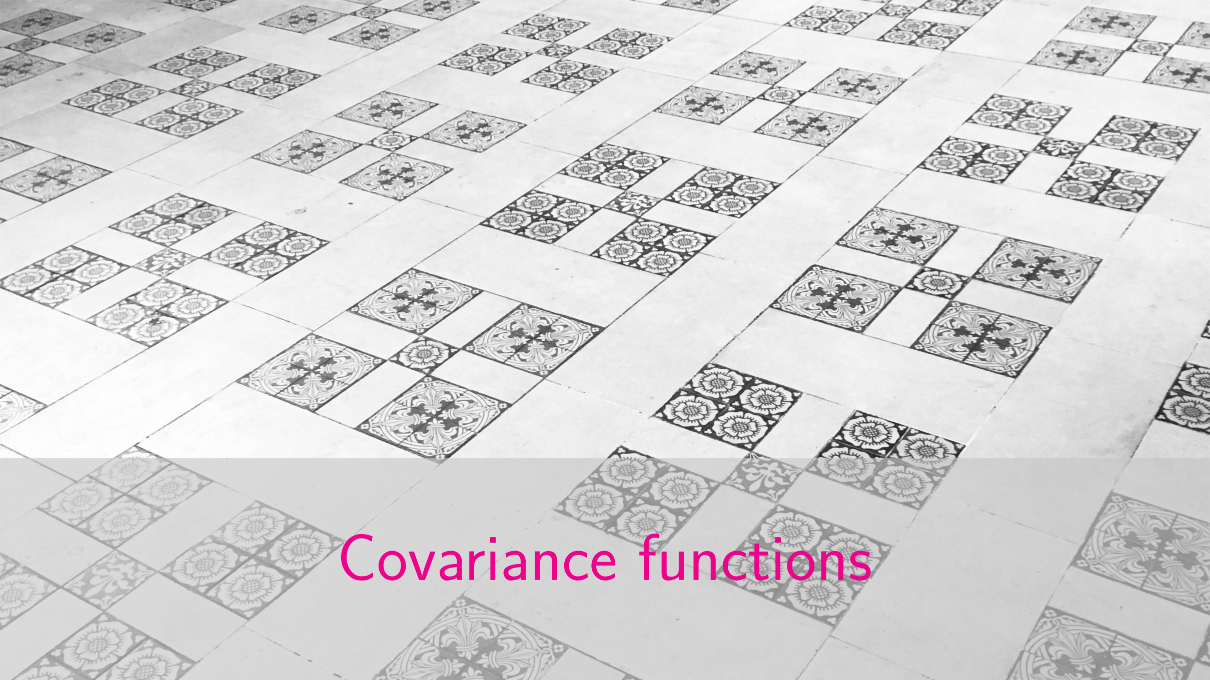


Why does Bayesian inference work?



Why does Bayesian inference work? Occam's Razor.





Covariance functions

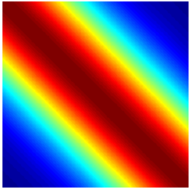
What effect does the form of the covariance function have?

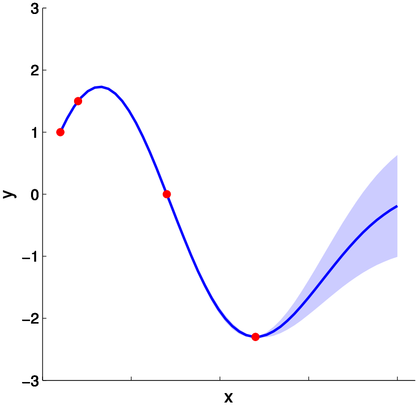
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Squared exponential

Exponentiated Quadratic

RBF covariance function

$\Sigma =$ 



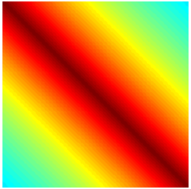
What effect does the form of the covariance function have?

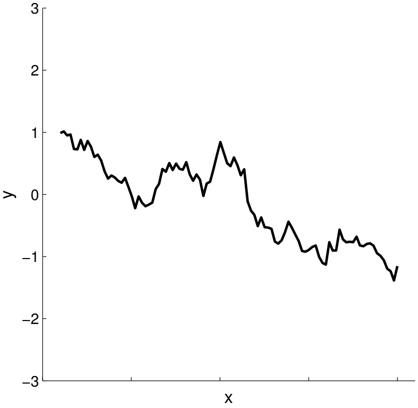
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



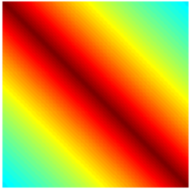
What effect does the form of the covariance function have?

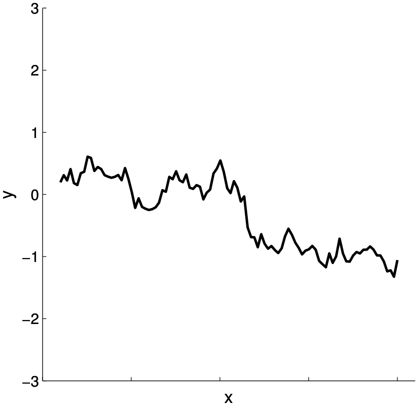
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



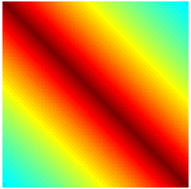
What effect does the form of the covariance function have?

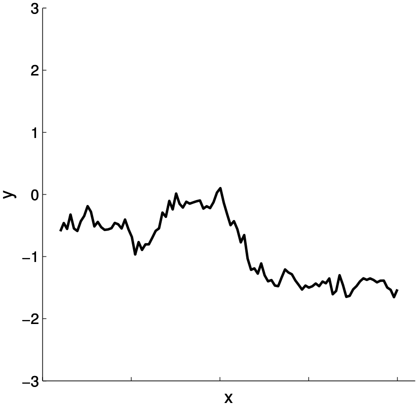
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

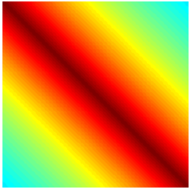
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

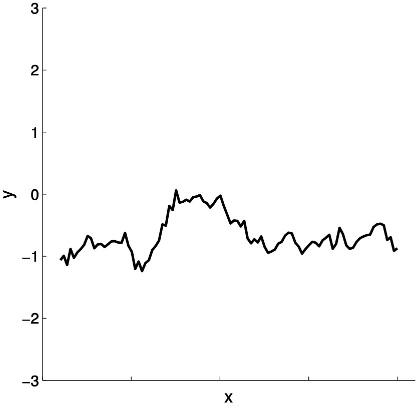
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

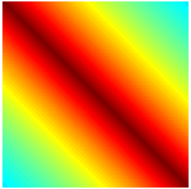
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

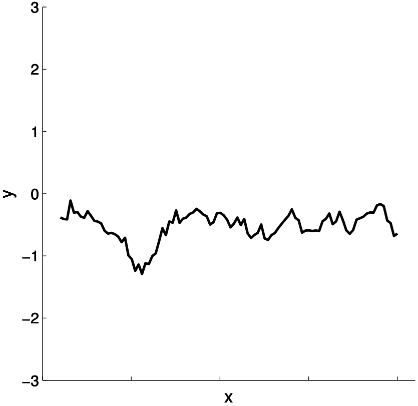
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

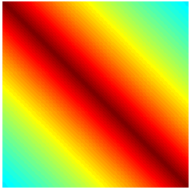
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

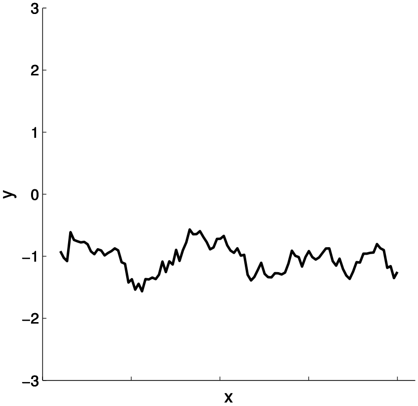
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





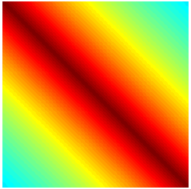
What effect does the form of the covariance function have?

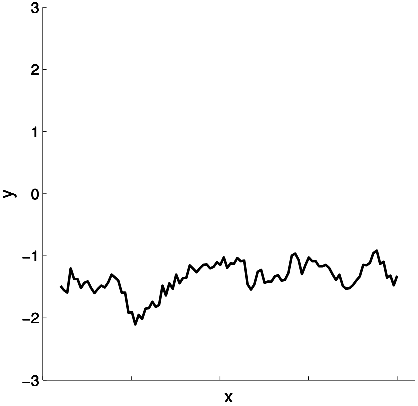
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

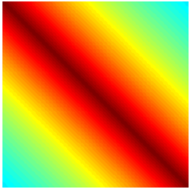
$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \exp\left(-\frac{1}{2\mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)$$

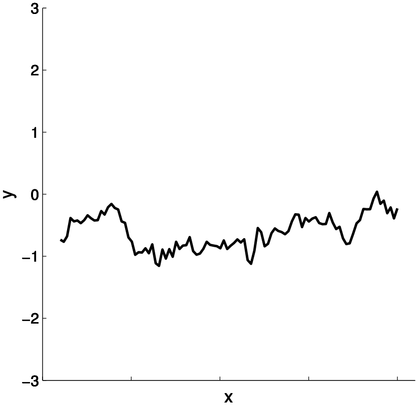
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

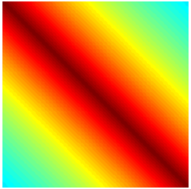
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

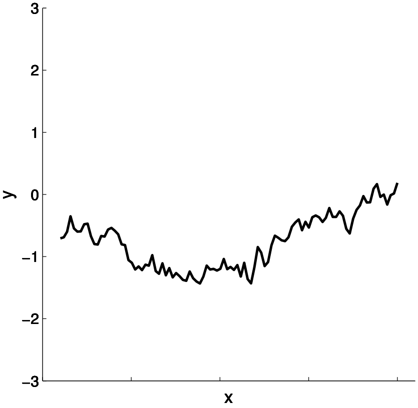
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





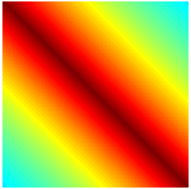
What effect does the form of the covariance function have?

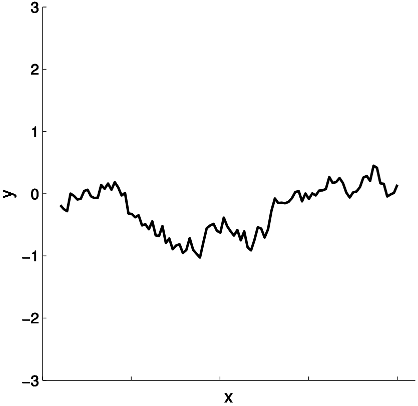
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



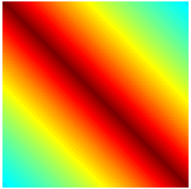
What effect does the form of the covariance function have?

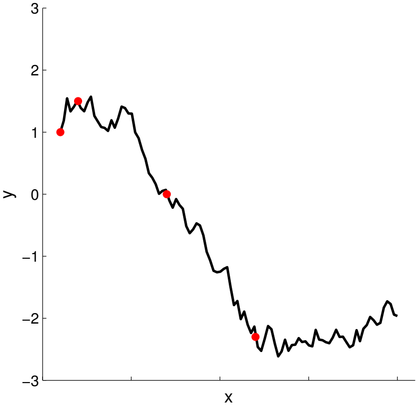
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



What effect does the form of the covariance function have?

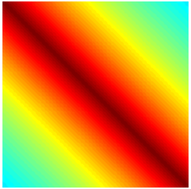
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

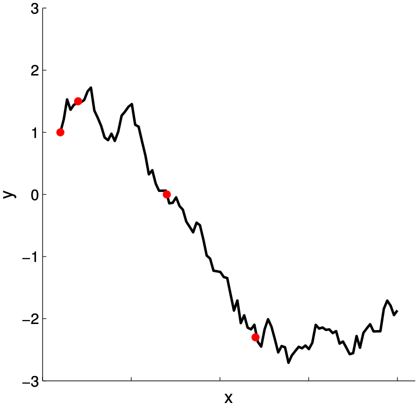
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





What effect does the form of the covariance function have?

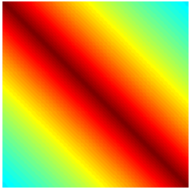
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

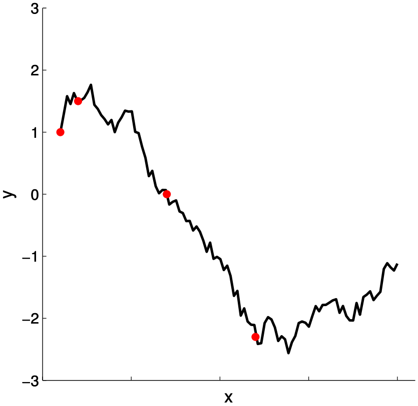
Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$





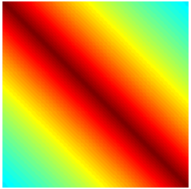
What effect does the form of the covariance function have?

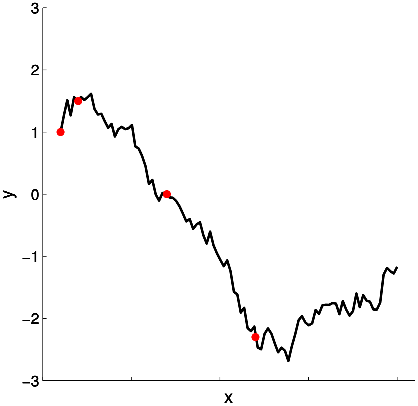
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



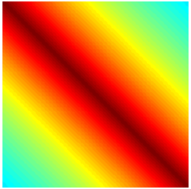
What effect does the form of the covariance function have?

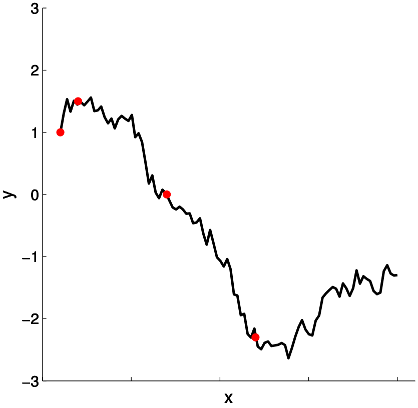
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



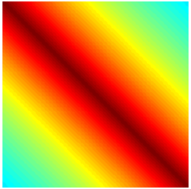
What effect does the form of the covariance function have?

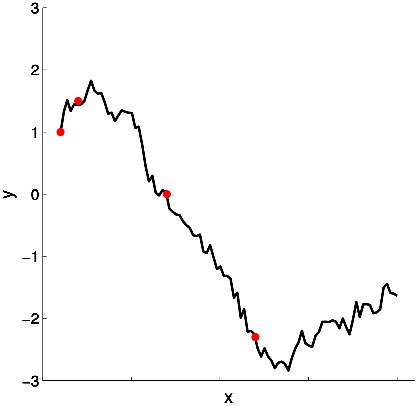
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



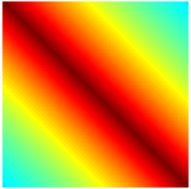
What effect does the form of the covariance function have?

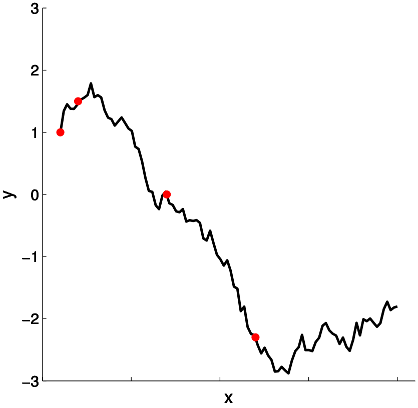
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



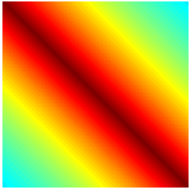
What effect does the form of the covariance function have?

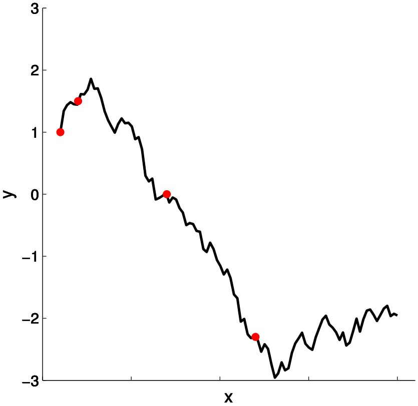
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



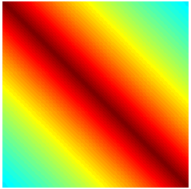
What effect does the form of the covariance function have?

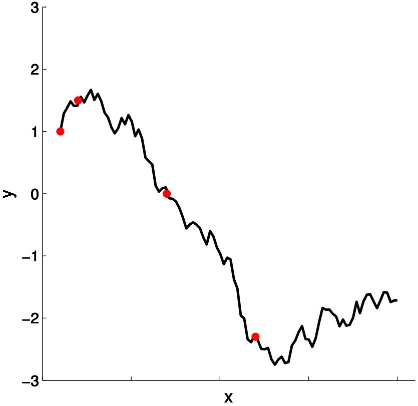
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



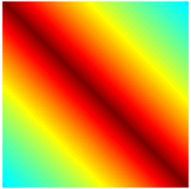
What effect does the form of the covariance function have?

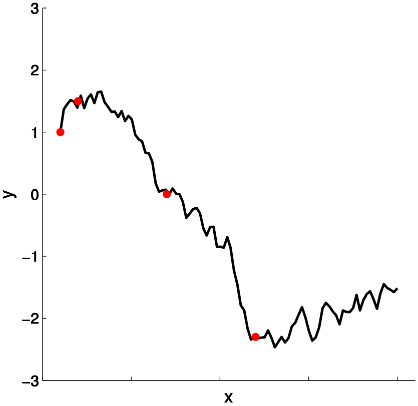
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

$\Sigma =$ 



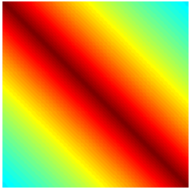
What effect does the form of the covariance function have?

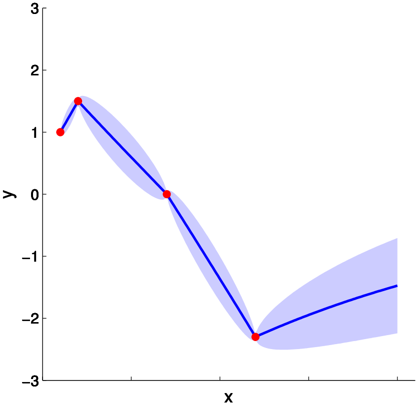
$$K(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2l^2}|x_1 - x_2|\right)$$

Laplacian covariance function

Browninan motion

Ornstein-Uhlenbeck

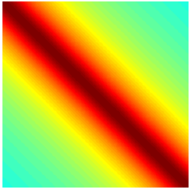
$\Sigma =$ 

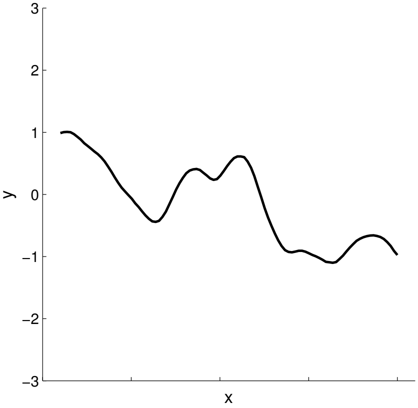


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

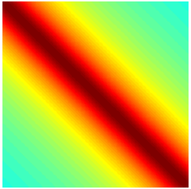
$\Sigma =$ 

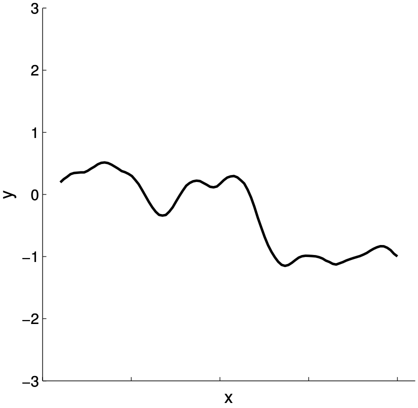


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

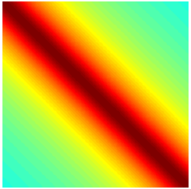
$\Sigma =$ 

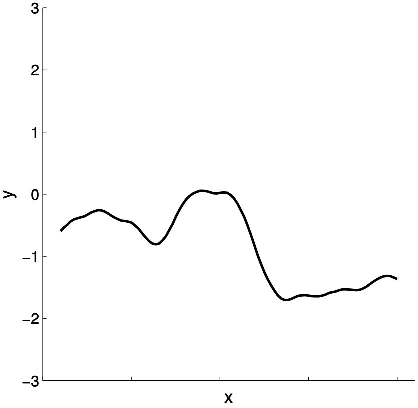


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

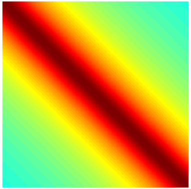


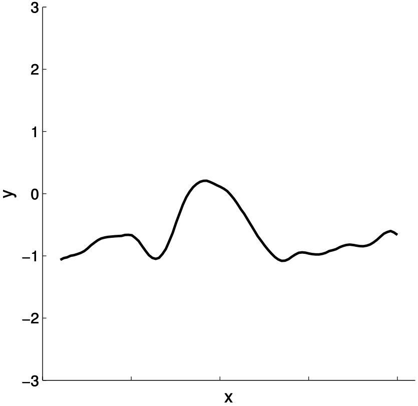
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$



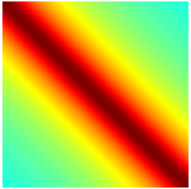


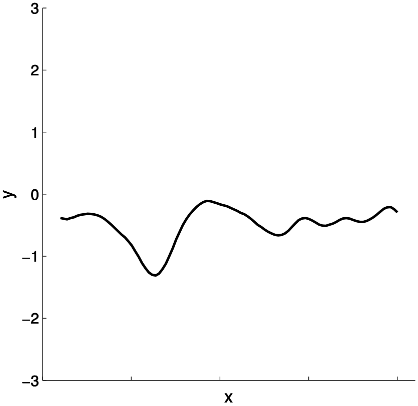
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

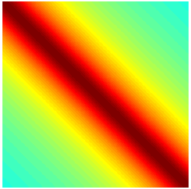


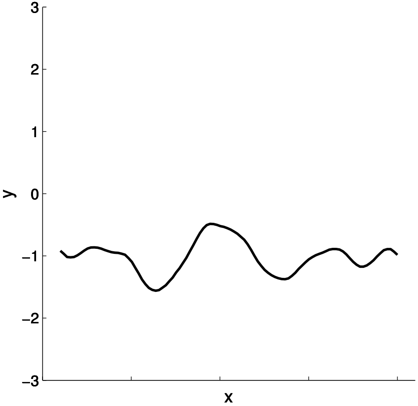


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

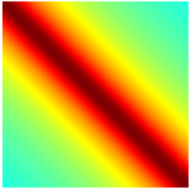


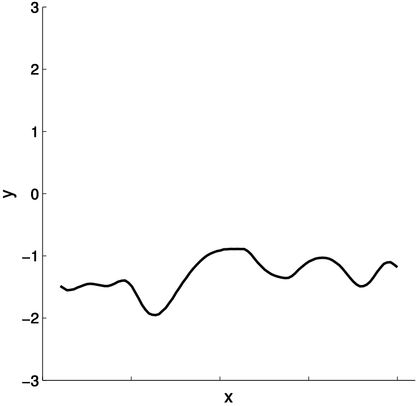
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$



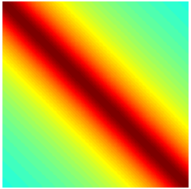


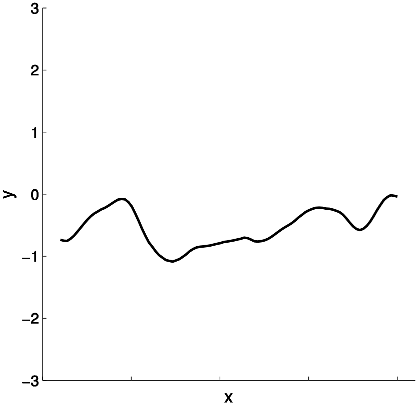
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

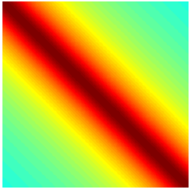


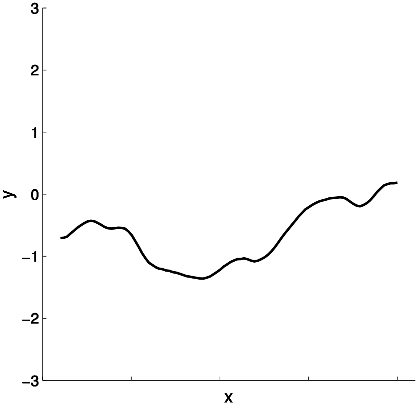


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2} |\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

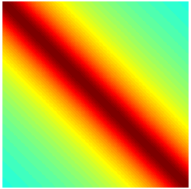


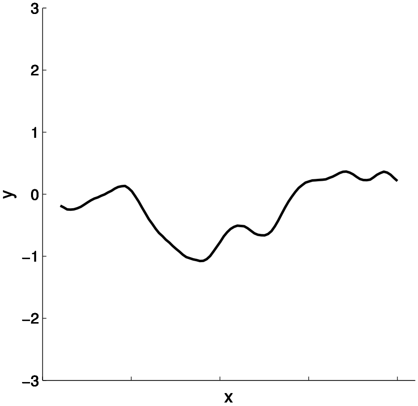
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

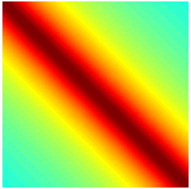


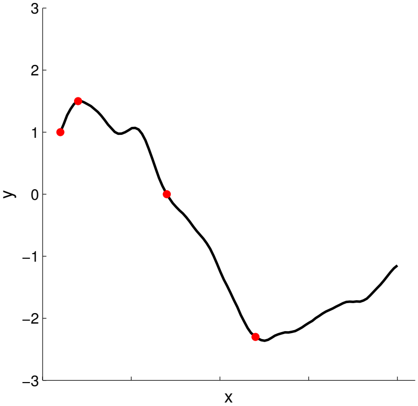


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

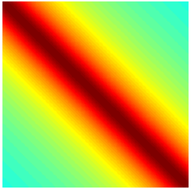


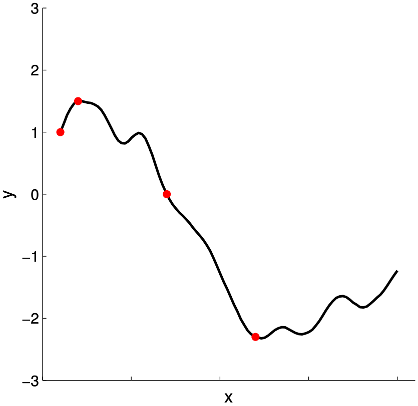
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$



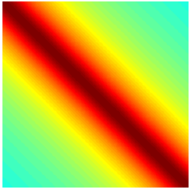


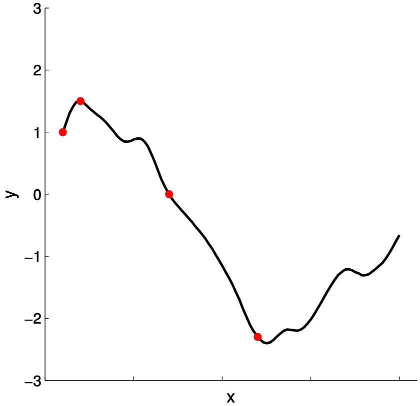
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

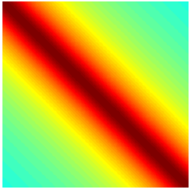


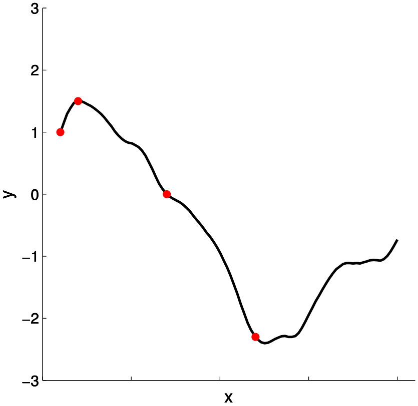


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

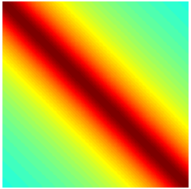
$\Sigma =$ 

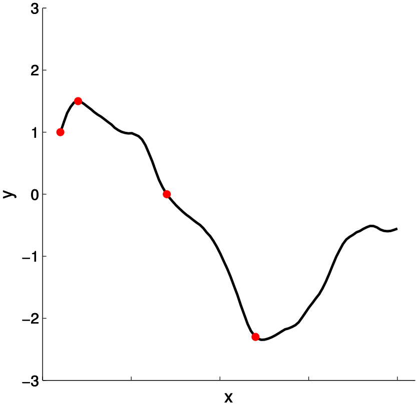


What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \left(1 + \frac{1}{2\alpha \mathsf{l}^2}|\mathsf{x}_1 - \mathsf{x}_2|\right)^{-\alpha}$$

Rational Quadratic

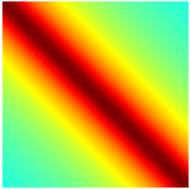
$\Sigma =$ 

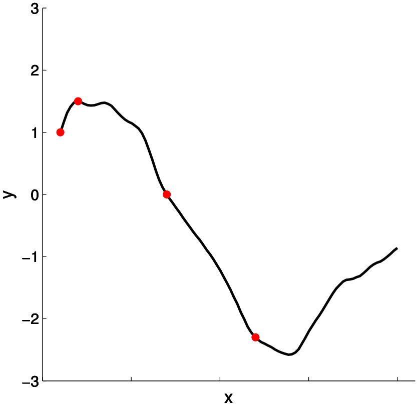


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

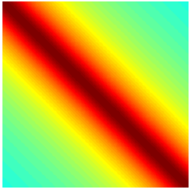
$\Sigma =$ 

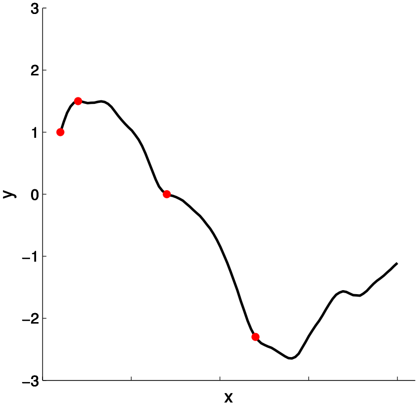


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

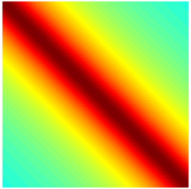
$\Sigma =$ 

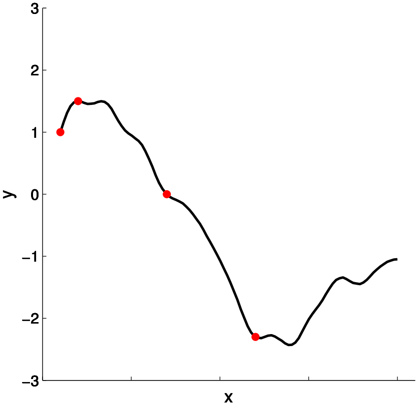


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

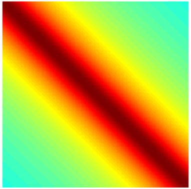
$\Sigma =$ 

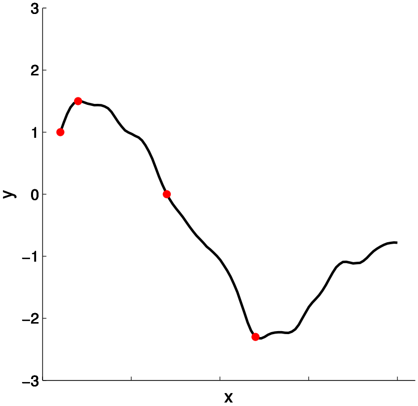


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2} |x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

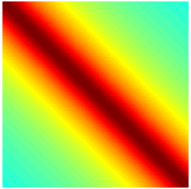


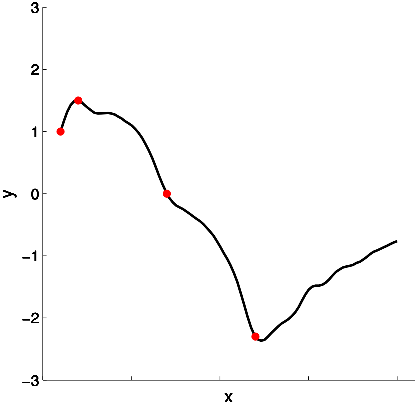
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2} |x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$

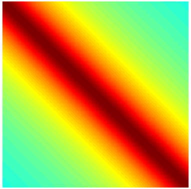


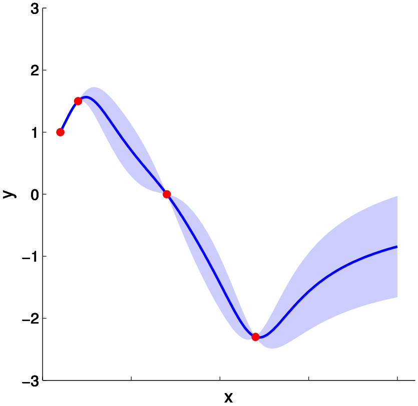


What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \left(1 + \frac{1}{2\alpha l^2}|x_1 - x_2|\right)^{-\alpha}$$

Rational Quadratic

$\Sigma =$ 

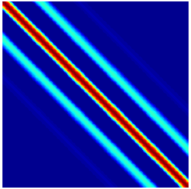


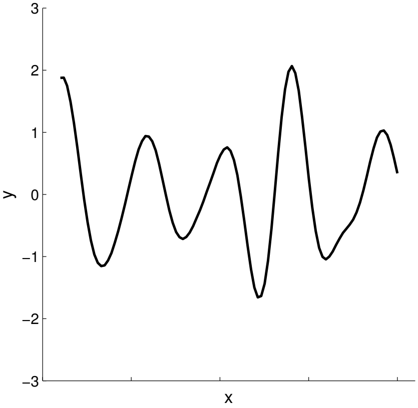
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 



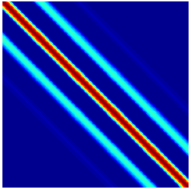
What effect does the form of the covariance function have?

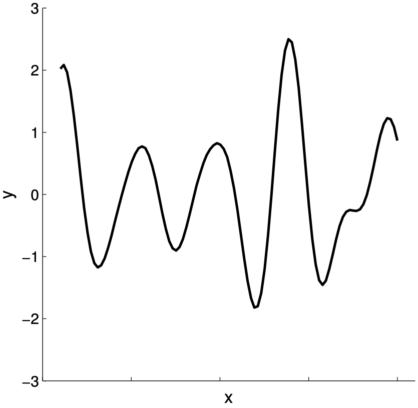
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



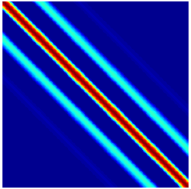


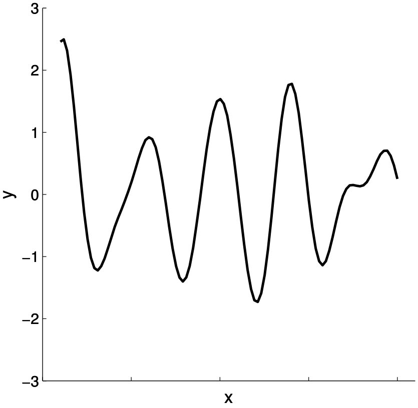
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

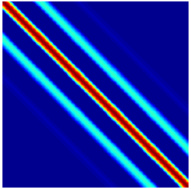


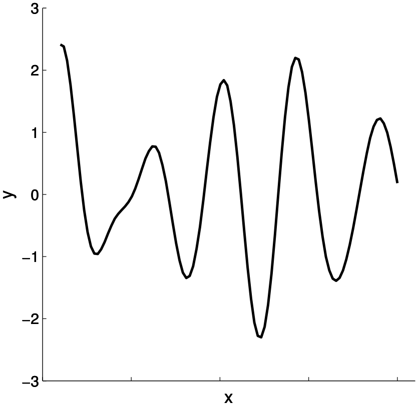
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 



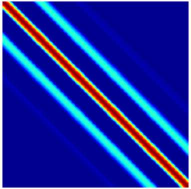
What effect does the form of the covariance function have?

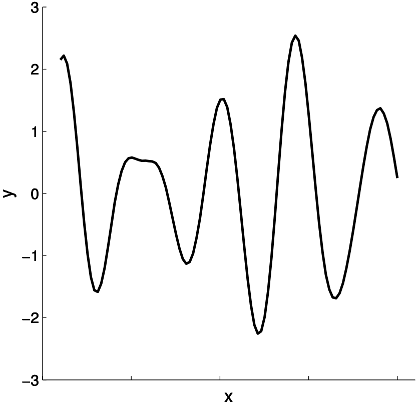
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



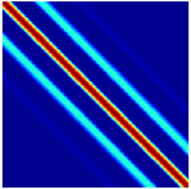


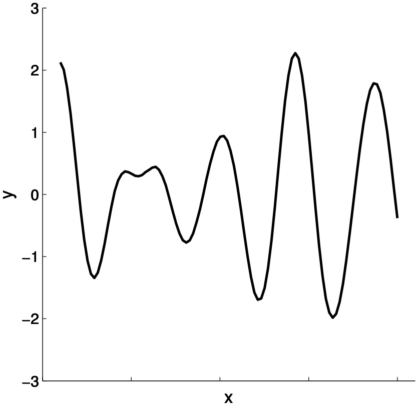
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

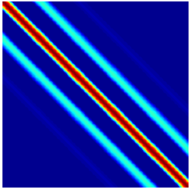


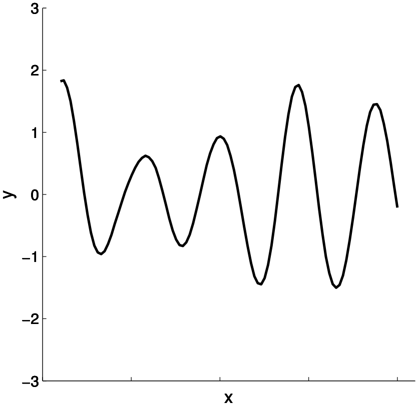
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

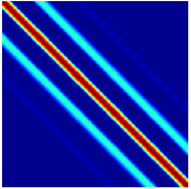


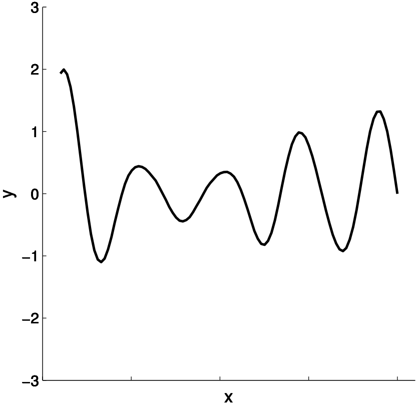
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid × squared exponential

$\Sigma =$ 

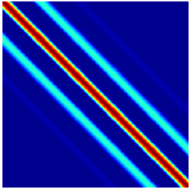


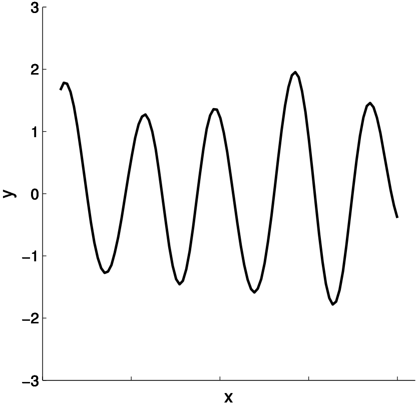
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

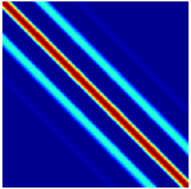


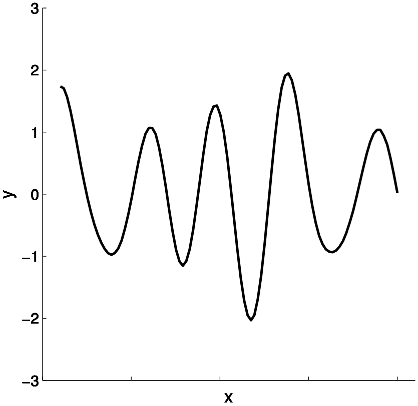
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

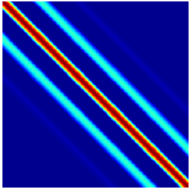


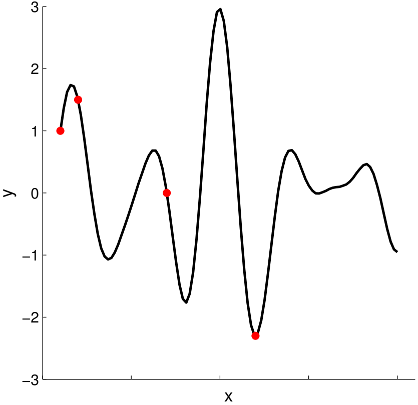
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 



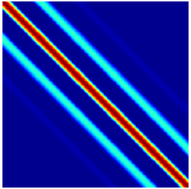
What effect does the form of the covariance function have?

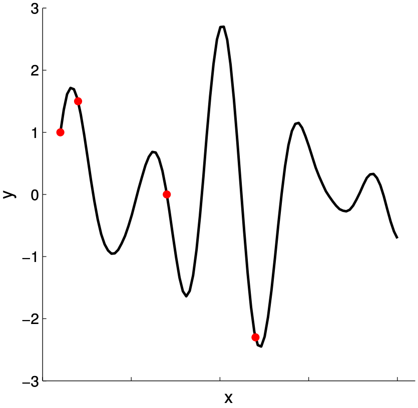
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



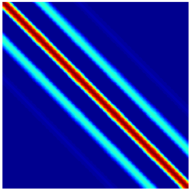


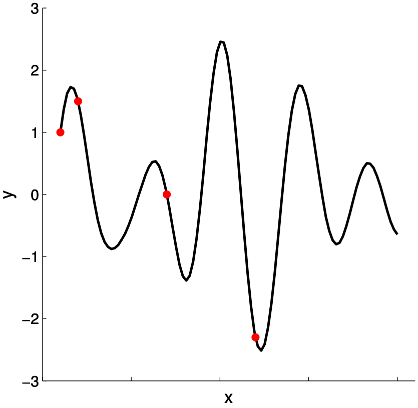
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid × squared exponential

$\Sigma =$ 

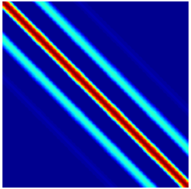


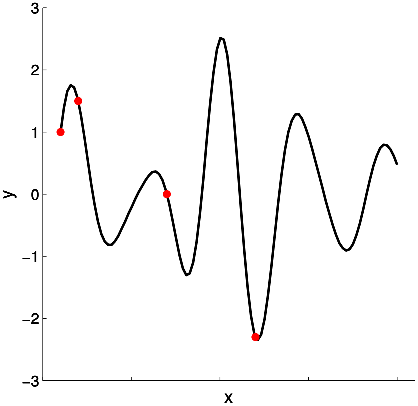
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

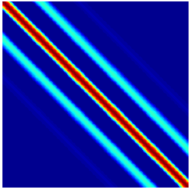


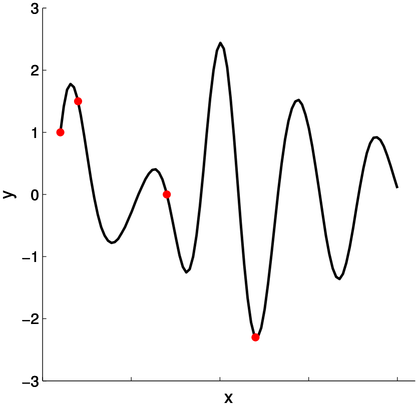
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

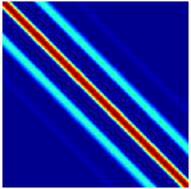


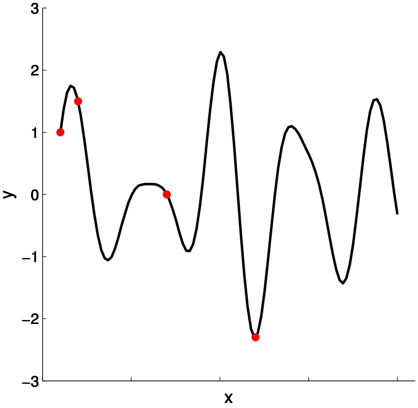
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

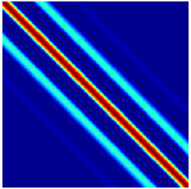


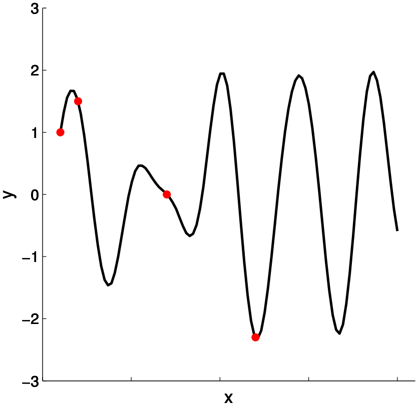
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 



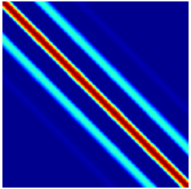
What effect does the form of the covariance function have?

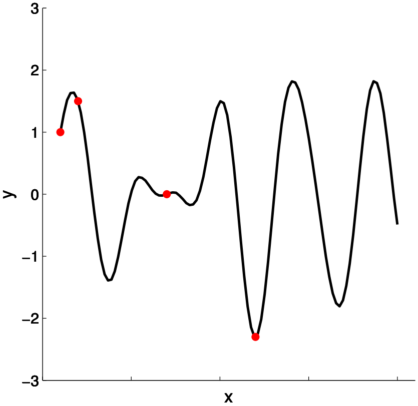
$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$



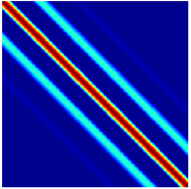


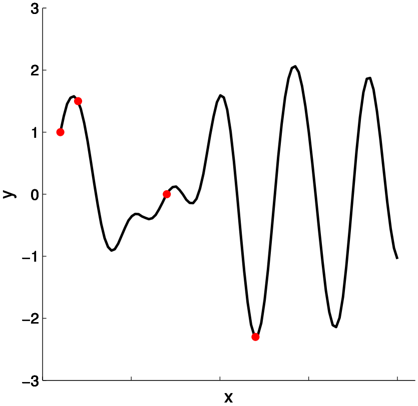
What effect does the form of the covariance function have?

$$\mathsf{K}(\mathsf{x}_1, \mathsf{x}_2) = \sigma^2 \cos(\omega(\mathsf{x}_1 - \mathsf{x}_2)) \exp\left(-\frac{1}{2\mathsf{l}^2}(\mathsf{x}_1 - \mathsf{x}_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

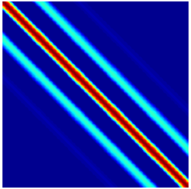


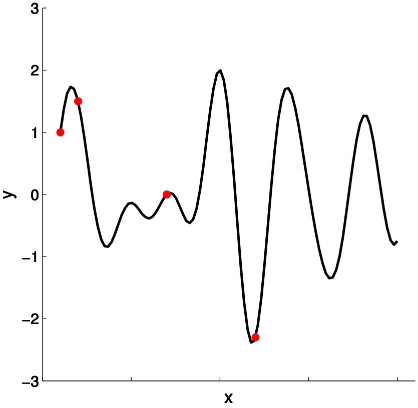
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

Periodic

sinusoid $\times$ squared exponential

$\Sigma =$ 

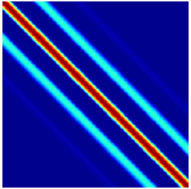


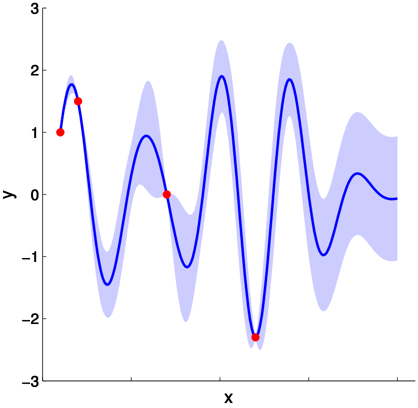
What effect does the form of the covariance function have?

$$K(x_1, x_2) = \sigma^2 \cos(\omega(x_1 - x_2)) \exp\left(-\frac{1}{2l^2}(x_1 - x_2)^2\right)$$

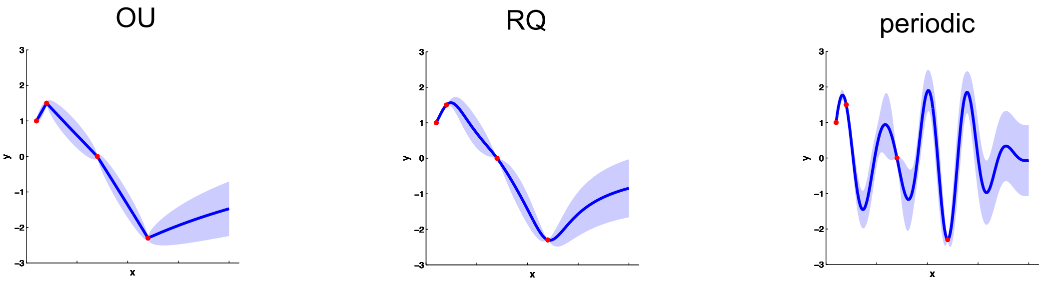
Periodic

sinusoid $\times$ squared exponential

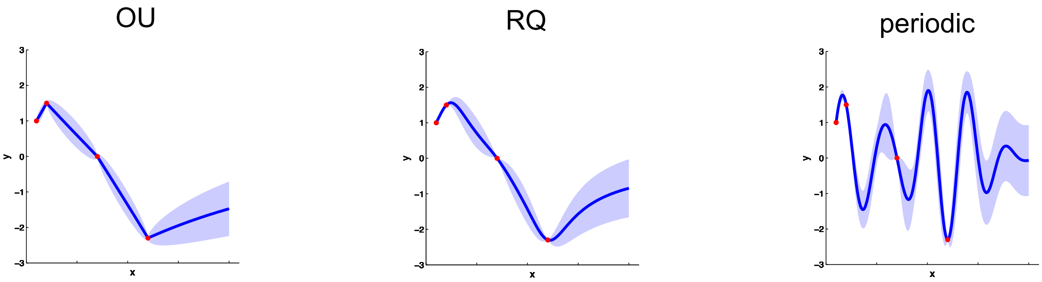
$\Sigma =$ 



The covariance function has a large effect



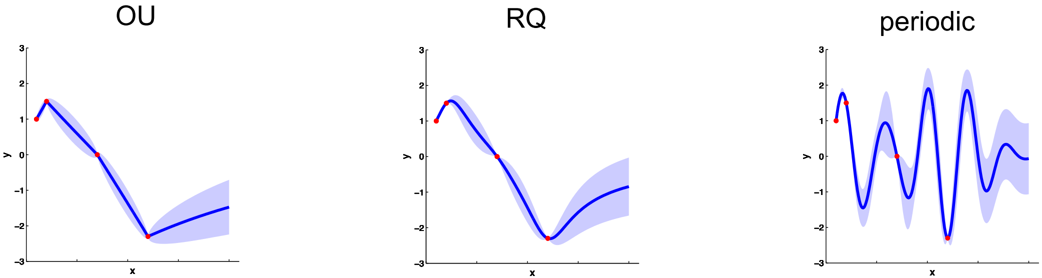
The covariance function has a large effect



Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

The covariance function has a large effect

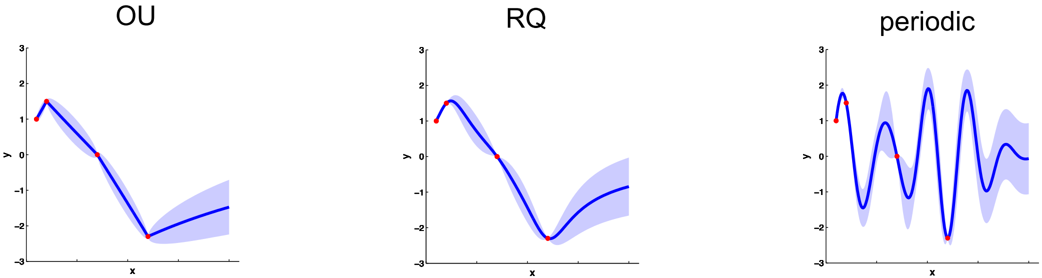


Bayesian model comparison:

$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

The covariance function has a large effect



Bayesian model comparison:

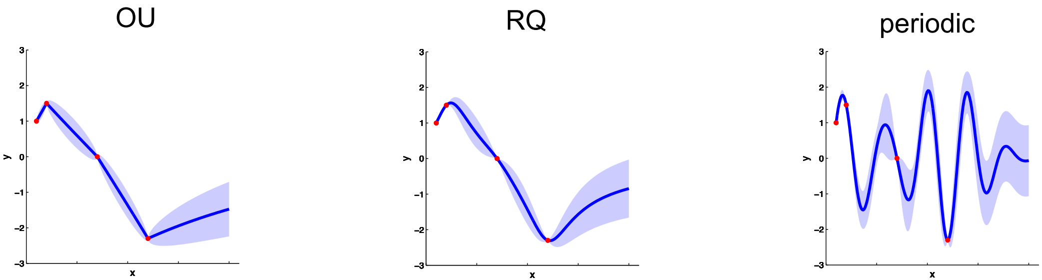
$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

marginal likelihood

$$p(\mathbf{y}_{1:N}|M) = \int \mathrm{d}\theta \; p(\mathbf{y}_{1:N}|\theta, M)p(\theta|M)$$

The covariance function has a large effect



Bayesian model comparison:

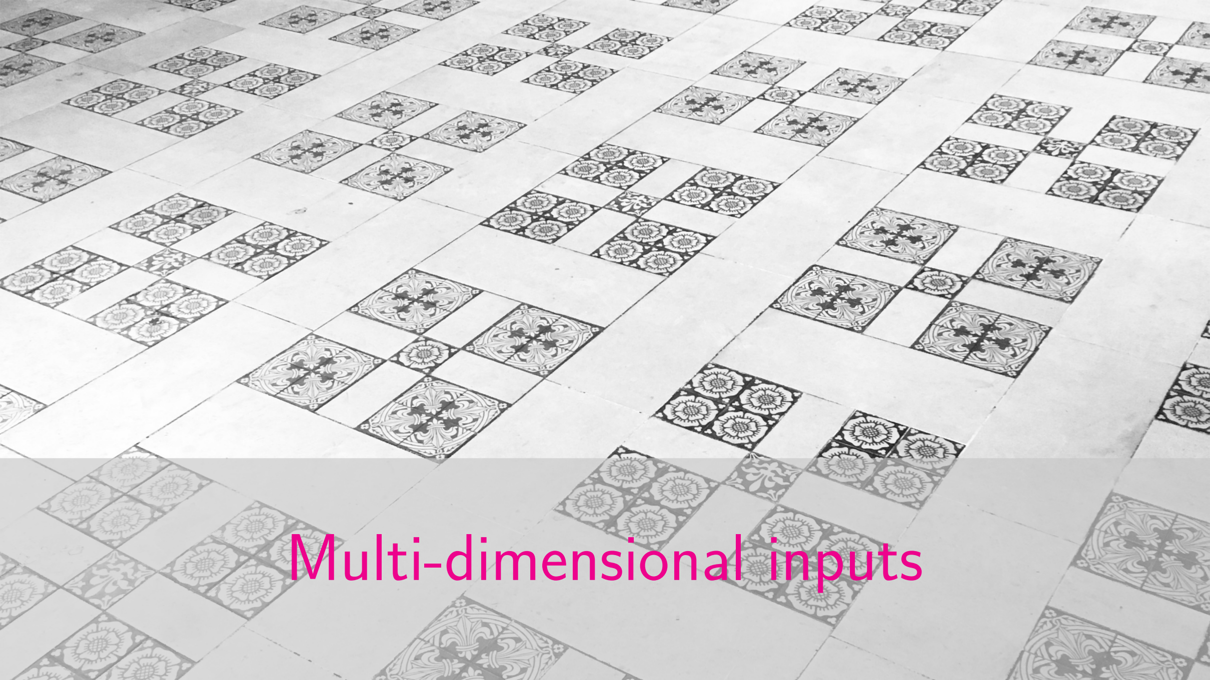
$$p(M|\mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N}|M)p(M)}{\sum_{M'} p(\mathbf{y}_{1:N}|M')p(M')}$$

prior over models

marginal likelihood

$$p(\mathbf{y}_{1:N}|M) = \int \mathrm{d}\theta \; p(\mathbf{y}_{1:N}|\theta, M)p(\theta|M)$$

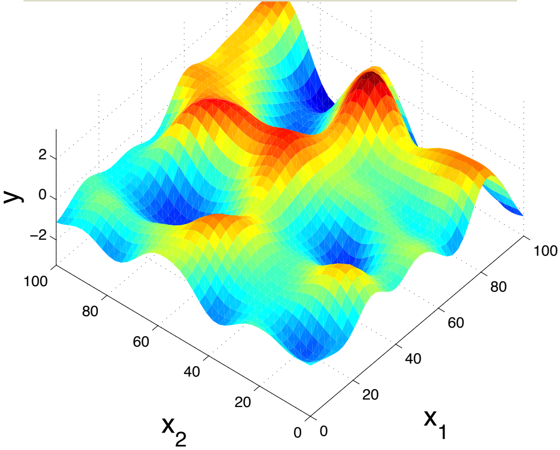
Health warnings:
Hard to compute (need approx.)
Results very sensitive to priors



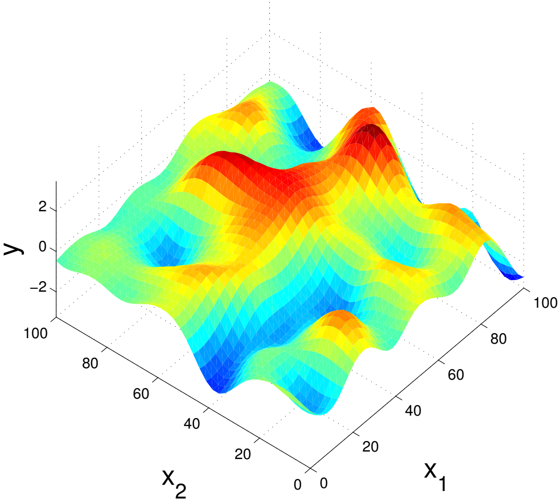
Multi-dimensional inputs

Higher dimensional input spaces

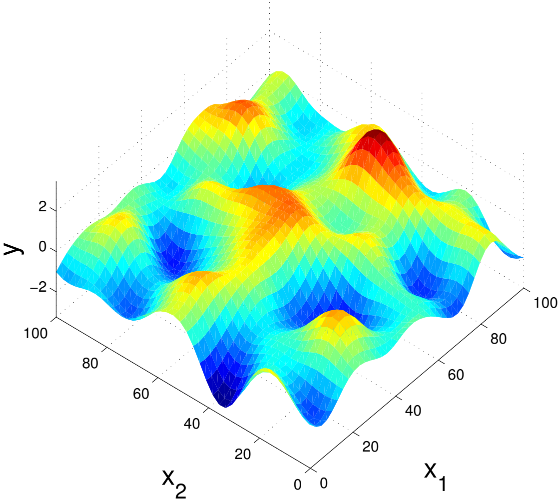
$$K(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2l_1^2}(\mathsf{x}_1 - \mathsf{x}_1')^2 - \frac{1}{2l_2^2}(\mathsf{x}_2 - \mathsf{x}_2')^2\right)$$



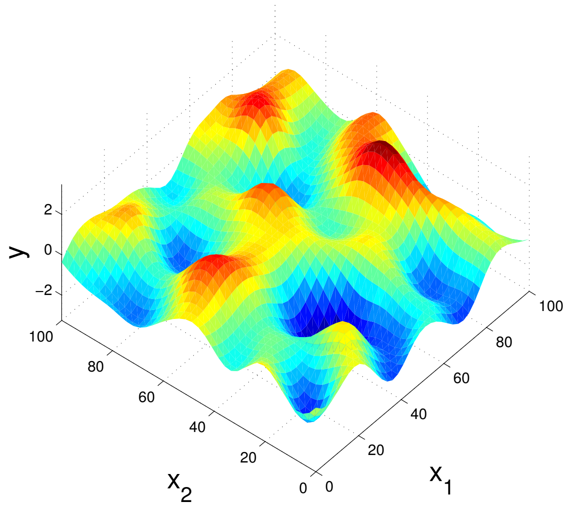
Higher dimensional input spaces



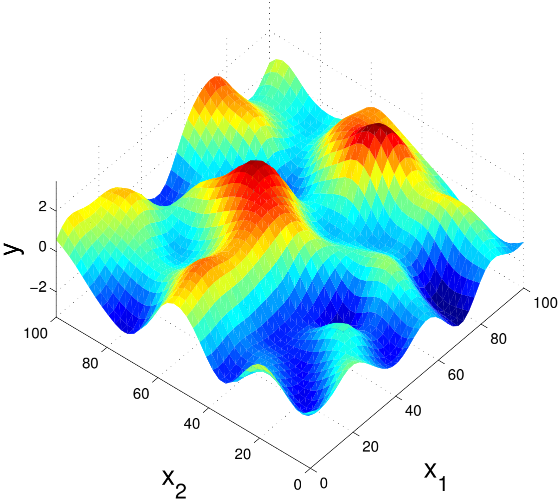
Higher dimensional input spaces



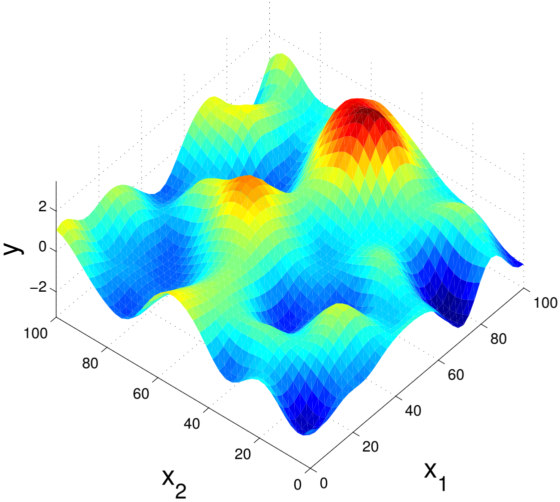
Higher dimensional input spaces



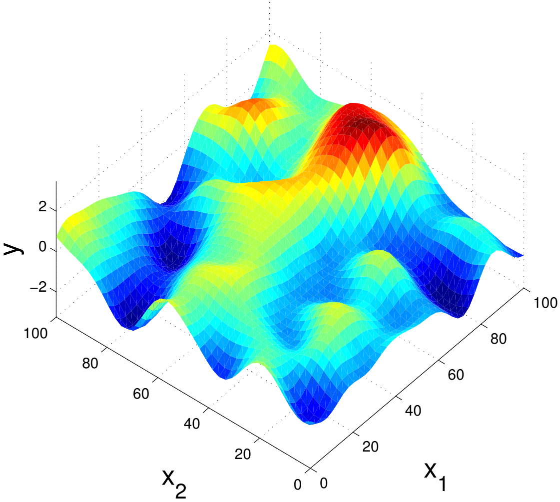
Higher dimensional input spaces



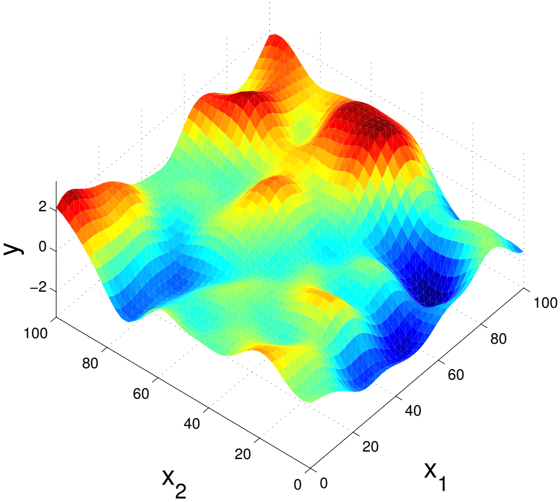
Higher dimensional input spaces



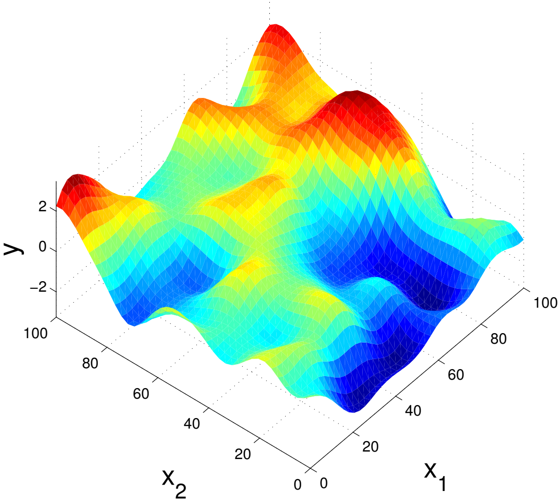
Higher dimensional input spaces



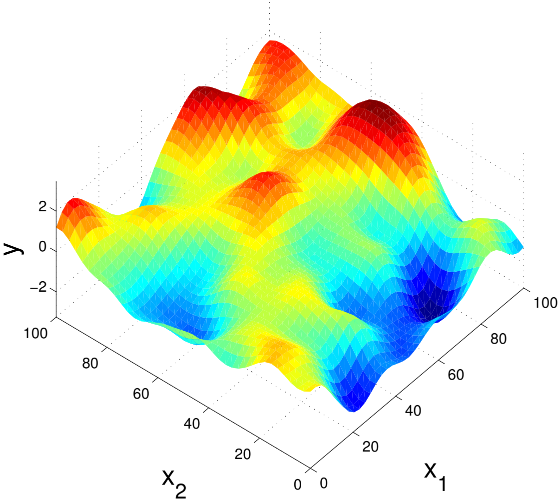
Higher dimensional input spaces



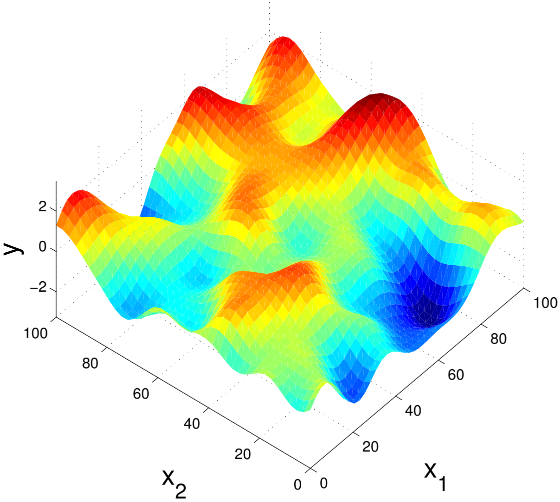
Higher dimensional input spaces



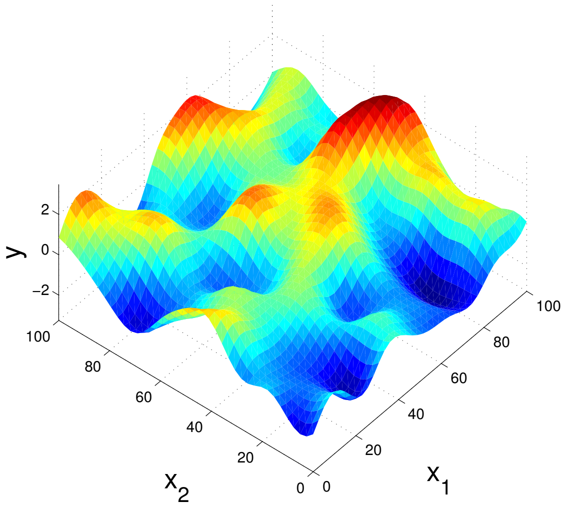
Higher dimensional input spaces



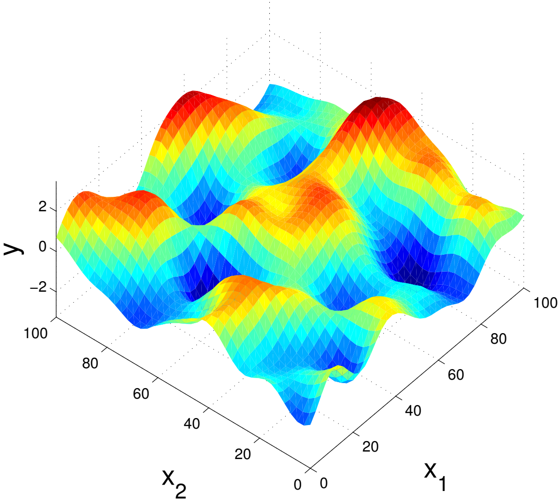
Higher dimensional input spaces



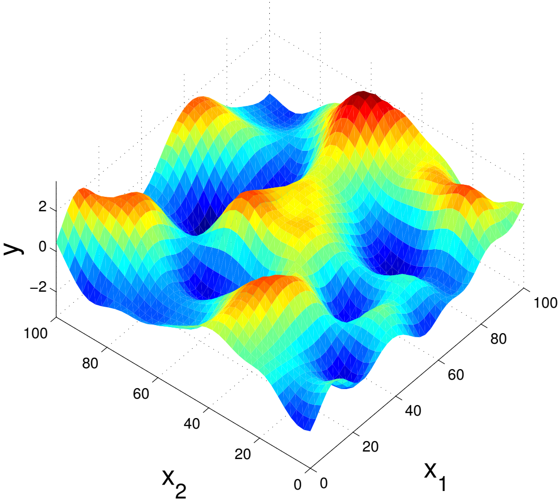
Higher dimensional input spaces



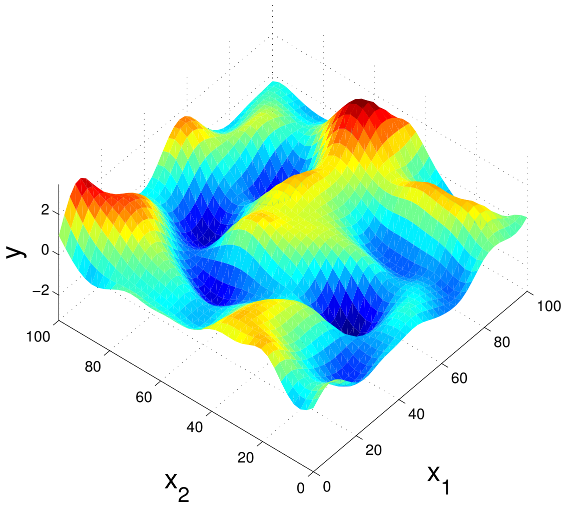
Higher dimensional input spaces



Higher dimensional input spaces



Higher dimensional input spaces



References (hyperlinked)

Great textbook available online:

- Gaussian Processes for Machine Learning, Rasmussen and Williams, 2006

Great Summer and Winter School:

- Gaussian Process Summer School, Neil Lawrence and colleagues

Software:

- GPy: Gaussian Processes in Python
- GPflow: Gaussian Processes and tensorflow
- GPML: Gaussian Processes in Matlab
- GP Stan: Gaussian Processes in probabilistic programming